Maize 16kD γ-zein forms very unusual disulfide-bonded polymers in the endoplasmic reticulum: implications for prolamin evolution

Davide Mainieri,[1] Claudia A. Marrano,[1] Bhakti Prinsi,[2] Dario Maffi,[2] Marc Tschofen,[3] Luca Espen,[2] Eva Stöger,[3] Franco Faoro,[2] Emanuela Pedrazzini,*[,1] Alessandro Vitale*[,1]

[1]Istituto di Biologia e Biotecnologia Agraria, CNR, 20133 Milano, Italy

[2]Dipartimento di Scienze Agrarie e Ambientali, Università degli Studi di Milano, 20133 Milano, Italy

[3]Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria

***Correspondence:** vitale@ibba.cnr.it tel: +39 02 23699 431; pedrazzini@ibba.cnr.it tel: +39 02 23699 443

E-mails of the other authors: Davide Mainieri mainieri@ibba.cnr.it; Claudia A. Marrano claudia.marrano@unimib.it; Bhakti Prinsi bhakti.prinsi@unimi.it; Dario Maffi dario.maffi@unimi.it; Marc Tschofen marc.tschofen@boku.ac.at; Luca Espen luca.espen@unimi.it; Eva Stöger eva.stoeger@boku.ac.at; Franco Faoro franco.faoro@unimi.it.

Running title: 16kD γ-zein forms unusual polymers in the ER

Highlight

A prolamin paralog generated upon maize whole genome duplication has changed its polymerization and solubility properties, allowing a new function in the assembly of maize protein bodies.

Abstract

In the lumen of the endoplasmic reticulum (ER), prolamin storage proteins of cereal seeds form very large ordered heteropolymers termed protein bodies (PBs), insoluble unless treated with alcohol or reducing agents. In maize PBs, 16kD γ-zein locates at the interface between a core of alcohol-soluble α-zeins and the outermost layer mainly composed of the reduced-soluble 27kD γ-zein. 16kD γ-zein originates from 27kD γ-zein upon whole genome duplication and is mainly characterized by deletions in the N-terminal domain that eliminate most Pro-rich repeats and part of the Cys residues involved in inter-chain bonds. 27kD γ-zein forms insoluble PBs also when expressed in transgenic vegetative tissues. We show that in Arabidopsis leaves 16kD γ-zein assembles into disulfide-linked polymers that fail to efficiently become insoluble. Instead of forming PBs, these polymers accumulate as very unusual threads that markedly enlarge the ER lumen, resembling amyloid-like fibers. Domain swapping between the two γ-zeins indicates that the N-terminal region of 16kD γ-zein has a dominant effect in preventing full insolubilization. Therefore, a newly evolved prolamin has lost the ability to form homotypic PBs, acquiring a new function in the assembly of natural, heteropolymeric PBs.

Keywords: cereal seeds, disulfide bonds, endoplasmic reticulum, genome wide duplication, neofunctionalization, prolamins, protein bodies, protein evolution.

Introduction

Prolamins are present only in the seeds of grasses, where they are usually the main proteins, thus constituting the major global source of food protein (Shewry and Halford, 2002). Their most striking and unique cell biology feature is their accumulation within the lumen of the

endoplasmic reticulum (ER) as very large heteropolymers termed protein bodies (PB, Shewry and Halford, 2002; Pedrazzini *et al.*, 2016). Most proteins that enter the ER are destined to secretion or distal locations of the endomembrane system, whereas ER residents, which are mainly folding helpers, have specific amino acid signals that allow their retention/retrieval in the ER (Gomez-Navarro and Miller, 2016). Since these signals are not present in prolamins, the question arises of which are the molecular features that determine prolamin ER residence and ordered PB formation.

Maize (*Zea mays*) prolamins are divided into four classes: α–zeins (more than thirty genes), three γ-zein genes, and the single δ and β genes (Woo *et al.*, 2001; Xu and Messing, 2008). 27kD γ–zein and β-zein are the oldest maize prolamins (Xu and Messing, 2008). Whole genome duplications (WGD), particularly common in plants (Jiao *et al.*, 2011), are followed by rearrangements that can lead to gene loss or retention. In the latter case, functional buffering or neofunctionalization can occur, playing important roles in evolution (Chapman *et al.*, 2006, Kassahn *et al.*, 2009). 5 to 12 million years ago, maize underwent WGD, followed by allotetraploidization (Swigoňová *et al.*, 2004). Thus γ-zein, originally a single gene encoding a polypeptide of 27kD and one of the most ancient maize prolamins, has now representatives in homologous regions of chromosome 7 (27 and 50kD γ-zein, hereon 27γz and 50γz) and 2 (16kD γ-zein, 16γz). Most probably, 16γz originates from 27γz gene duplication followed by deletion events (Xu and Messing, 2008).

During endosperm development, γ– and β-zeins are synthesized first, forming a PB where α– and δ-zein will later accumulate (Lending and Larkins, 1989). In the mature PB, β-zein, 27γz and 50γz form the outer layer in contact with the luminal face of the ER membrane, whereas α– and δ-zein form the inner core, with 16γz located at the interface between the core and the outer layer (Lending and Larkins, 1989; Yao *et al.*, 2016). Yeast two-hybrid data suggest that 16γz can interact with zeins of all classes (Kim *et al.*, 2002, 2006). 27γz expressed in vegetative tissues of transgenic plants forms homotypic PBs, indicating that no specific features of the maize endosperm ER are necessary to form a PB (Geli *et al.*, 1994). The primary sequence of 27γz (Fig.1) consists of the transient signal peptide for translocation into the ER (co-translationally removed), followed by a region containing eight or seven (depending on the maize variety) repeats of the hexapeptide PPPVHL and seven Cys residues involved in inter-chain bonds that make the protein insoluble in non-reducing conditions, and finally a second region homologous to 2S albumins, which are vacuolar storage proteins present in various amounts in all land plants (Vitale *et al.*, 1982; Prat *et al.*, 1985; Mainieri *et al.*, 2014). 2S albumins belong to a larger class characterized by the eight cysteine motif, consisting in four

intrachain disulfide bonds between three helical domains (Pedrazzini *et al.*, 2016; Fig. 1). This motif is also conserved in 27γz (Ems-McClung *et al.*, 2002). Progressive Cys to Ser mutation of the seven Cys residues of the N-terminal region led to increased solubility and parallel increase in the ability to leave the ER along the secretory pathway (Mainieri *et al.*, 2014). When the N-terminal region including the first six Cys residues was fused at the C-terminus of phaseolin, the vacuolar 7S storage globulin of common bean, the chimeric protein zeolin formed homotypic PBs in the ER (Mainieri *et al.*, 2004). Zeolin was instead efficiently secreted upon in vivo treatment with reducing agent, or when its six Cys residues were mutated to Ser (Pompa and Vitale, 2006). Overall, these studies indicate that the N-terminal region of 27γz contains key information for PB assembly and that its Cys residues are necessary for this process.

16γz (Fig. 1) is mainly characterized by the loss of large part of the N-terminal, Pro-rich domain and three of its seven Cys residues (Prat *et al.*, 1987). Additionally, its C-terminal region has lost one Cys residue of the eight-cysteine motif and has acquired a new one near the C-terminus, resulting in a new CysProCys sequence. This tripeptide could form an intra-chain disulfide bond (Yu *et al.*, 2012), however it is not known whether this occurs in 16γz. The changes that have generated 16γz are noteworthy, since Cys residues are rarely lost once acquired during evolution (Feyertag and Alvarez-Ponce, 2017; Wong *et al.*, 2011). 16γz can thus provide information on the minimal requirements for PB biogenesis and the features that allow the formation of heteropolymeric maize PBs. Here we show that, unlike 27γz, ectopically expressed 16γz remains in part soluble, mainly because of the mutations in the N-terminal region. 16γz is unable to form PBs, but it stably accumulates as polymers that markedly enlarge the ER lumen, giving rise to very unusual filamentous structures. These characteristics indicate neofunctionalization after WGD and cast light on the molecular basis for the specific organization of maize PBs.

Materials and Methods

Analysis of maize PBs

Seeds from *Zea mays* inbred line W64A, collected at 25 days post-pollination and stored at -80 °C, were homogenized in a mortar using 5 ml g^{-1} ice-cold 100mM Tris-Cl, pH 7.4, 1.0 mM EDTA (buffer H), 7% (w/w) sucrose, Complete protease inhibitor cocktail (Roche). After cheesecloth filtration, the homogenate was loaded on two layers of 35% and 60% (w/w) sucrose

in buffer H and centrifuged in a swinging rotor for 90 min, 4 °C at 78,900 g_{av}. The 7% sucrose supernatant, the interface between 7% and 35% sucrose and the interface between 35%, and 60%sucrose were collected. After denaturation in the presence of 1% SDS and 4% 2-mercaptoethanol (2-ME), proteins were analyzed by 15% SDS-PAGE. As expected (Vitale *et al.*, 1982), zeins were at the interface between 35% and 60% sucrose, hence termed PB fraction. To treat PBs with different solvents, immediately after collection the PB fraction was first diluted with the same volume of buffer H and centrifuged 10 min, 4°C, 1,500 g_{av}. The PB pellet was then resuspended in either of the following solvents: i) buffer H, 1% Triton X-100, 20 min, 4°C; ii) buffer H, 2mM dithiothreitol (DTT), 20 min, 4°C; iii) buffer H, 4% 2-ME, 20 min 4°C; iv) 70% ethanol in H_2O, 90 min, either at 4°C or 25°C. After each treatment, samples were centrifuged 10 min, 4°C, 1,500 g_{av}; pellet and supernatant were denatured and analyzed by 15% SDS-PAGE and staining with Coomassie Brilliant Blue. Protein Molecular Weight Markers (Fermentas, Vilnius, Lithuania) were used as molecular mass markers.

Plasmid constructions

The pDHA vector containing the coding sequence of 27kD γ-zein followed by the FLAG epitope DYKDDDDK (here named 27γzf) has been described (Mainieri *et al.*, 2014). To construct a similarly tagged 16 kD γ-zein (16γzf), a genomic fragment of *Zea mays* W64A comprising the coding and untranslated flanking sequences of 16kD γ-zein (a kind gift from Angelo Viotti), identical to GenBank sequence EU953296.1, was amplified by PCR using the following oligonucleotides:

5'- ACTCAGGTCGAC**ATG**AAGGTGCTGATCGTTGCCCTTG -3' (where the SalI restriction site is underlined and the 16kD γ-zein ATG start codon is in bold) and 5'-TCGATGGCATGCTCACTTGTCGTCGTCGTCCTTGTAGTCGTAGTAGACACCGCCGGCAGC -3', (where the SphI restriction site is underlined and the reverse complement of the codons encoding the FLAG epitope is double underlined). The sequence was restricted with SalI and SphI and reinserted into the similarly restricted pDHA vector, for transient expression.

To produce transgenic Arabidopsis, the EcoRI fragments containing the 16γzf or 27γzf expression cassettes were excised from pDHA and subcloned into EcoRI-linearized pGreenII0179 (http://www.pgreen.ac.uk, John Innes Centre, Norwich, Norfolk, UK). The *Agrobacterium tumefaciens* strain GV3101, containing the pSoup helper plasmid, was transformed with the resulting constructs.

To prepare the chimeric construct 16/27, which is formed by the N-terminal primary sequence of 16γzf until Cys^{73} followed by the C-terminal sequence of 27γzf starting from Gly^{118} (see also Fig. 1), DNA was synthesized (Integrated DNA Technologies, Leuven, Belgium) based on the two sequences and inserted into SalI/SphI restricted pDHA. To prepare the exactly reciprocal construct 27/16, the following DNA sequence was synthesized: from the Bpu10I restriction site of the sequence encoding 27γzf until its Gly^{118} codon (which corresponds to Gly^{74} of 16γzf), continuing with the 16γzf sequence from Val^{75} until its stop codon, and ending with a SphI restriction site. This sequence was used to substitute the Bpu10I/Sph1 fragment in pDHA encoding 27γzf.

Transient expression in tobacco protoplasts

Transient expression was performed in protoplasts prepared from young (4-7 cm) leaves of tobacco (*Nicotiana tabacum* SR1) grown in axenic conditions, as described (Mainieri *et al.*, 2014). Resuspensions of 10^6 protoplasts were transfected using 40 μg/million protoplasts of plasmid, or, for co-transfections, 60 μg (25 μg of each plasmid plus empty pDHA to final 60μg). After transfection and incubation for 20 h at 25°C, protoplasts were either homogenized for protein blot analysis or subjected to pulse-chase labelling. Extraction of intracellular and secreted proteins in reducing or oxidizing conditions and protein blot analysis with rabbit anti-FLAG antibody (1:2,000 dilution, Sigma-Aldrich) and the Super-Signal West Pico Chemiluminescent Substrate (Pierce Chemical, Rockford, IL) were performed as described (Mainieri *et al.*, 2014). Protein Molecular Weight Markers (Fermentas, Vilnius, Lithuania) were used as molecular mass markers.

Pulse-chase labelling was performed with 100 μCi ml^{-1} Easytag mixture of ^{35}S-labelled Met and Cys (PerkinElmer) for 1 h at 25°C. Chase was allowed by adding unlabeled Met and Cys to 10 mM and 5 mM, respectively. After incubation at 25°C for the desired chase time, 2 volume of ice-cold W5 buffer (Mainieri *et al.*, 2014) were added to each sample, which was then centrifuged at 60 *g,* 10 min. Collected protoplasts and supernatant (containing secreted proteins) were homogenized with 2 volumes of ice-cold 150 mM NaCl, 1.5 mM EDTA, 1.5 % Triton X-100, 150 mM Tris-Cl pH 7.5, supplemented with Complete. After centrifugation at 10,000 *g*, the pellet was resuspended in the same buffer supplemented with 4% 2-ME and centrifuged again. The soluble fractions of the first and second centrifugation were immunoselected using the anti-FLAG antibody and protein A Sepharose (GE Healthcare) and analyzed by SDS-PAGE and radiography, using ^{14}C-methylated proteins (Sigma-Aldrich) as

molecular mass markers. Radioactive proteins were detected using the Starion FLA-9000 Phosphoimage System (Fujifilm) and quantified using TotalLab Quant (TotalLab, Newcastle upon Tyne, UK).

Expression in transgenic Arabidopsis

Transgenic *Arabidopsis thaliana* (ecotype Columbia) plants expressing 16γzf or 27γzf were produced by floral dip with the transformed *Agrobacterium tumefaciens* described above. Hygromycin-resistant T0 plants were identified and the homozygous progenies were selected. Experiments were then conducted using T2 or T3 plants. Plants were grown in soil at 23°C under a 16/8 h light/dark cycle or in sterile conditions on half-concentrated Murashige and Skoog media (Duchefa Biochemie) supplemented with 10 g/L Sucrose and 0.8% (w/v) phyto agar (Duchefa Biochemie).

Four to six weeks old leaves were homogenized in leaf homogenization buffer (150 mM Tris-Cl, pH 7.5, 150 mM NaCl, 1.5 mM EDTA, 1.5% Triton X-100, Complete protease inhibitor cocktail [Roche]), supplemented (reducing conditions) or not (oxidizing conditions) with 4% (v/v) 2-ME. Soluble and insoluble proteins were separated by centrifugation at 1,500 *g*, 10 min, 4°C. Samples were adjusted to 1.0% SDS, 4% 2-ME and analyzed by SDS-PAGE followed by protein blot with the anti-FLAG antibody (1:2,000 dilution).

Subcellular fractionation

Four to six weeks old Arabidopsis leaves were homogenized in 10 mM KCl, 2 mM MgCl2, 100 mM Tris-Cl, pH 7.8 (buffer A), 12% (w/w) sucrose, 4°C, followed by isopycnic ultracentrifugation using linear 16-65% (w/w) sucrose gradients in buffer A as described (Mainieri *et al.*, 2004). Fractions of 650 µL were collected; 40 µL of each fraction were denatured and analysed by SDS-PAGE, followed by protein blot with anti-FLAG antibody or rabbit anti-endoplasmin serum (Klein *et al.*, 2006; 1:2,500 dilution).

To determine the solubility of γ-zeins present in the different subcellular fractions, fractions around either 1.19 or 1.29 density were frozen to break membranes and then pooled. Equal volume of buffer A was added and the suspension centrifuged 1,500 $g_{av,}$ 10 min, 4°C. Supernatants (soluble proteins) were collected and denatured with SDS-PAGE denaturation buffer. Pellets were either resuspended in SDS-PAGE denaturation buffer or were further extracted with 70% ethanol in H_2O, 90 min at 25°C and centrifuged 1,500 g_{av}, 10 min, 25 °C.

Soluble fraction (ethanol-soluble) and insoluble pellet (ethanol-insoluble) were collected and denatured for SDS-PAGE.

Velocity sucrose gradient ultracentrifugation

Arabidopsis leaves were homogenized in ice-cold leaf homogenation buffer. The homogenate was loaded on top of a linear sucrose gradient (150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 50 mM Tris-Cl, pH 7.5, 5% to 25% [w/v] sucrose). After centrifugation at 200,000 g_{av}, 4°C for 20 h, equal volumes of each fraction were analyzed by SDS-PAGE and protein blot. An identical gradient loaded with molecular mass markers was run in parallel. For velocity ultracentrifugation in reducing conditions, leaf homogenization buffer was supplemented with 4% 2-ME, and the sucrose gradient buffer was supplemented with 2% DTT.

Electron microscopy

Tissue fragments (1-2 mm^2) from Arabidopsis fully expanded leaves were fixed, embedded and immunolabelled as previously described (Faoro *et al.*, 1991). Tissues were fixed in 1.2% glutaraldehyde and 3.3% paraformaldehyde in 0.1 M phosphate buffer pH 7.4 at 4°C for 2 h, postfixed in 1% OsO_4 in the same buffer for 2 h, dehydrated in an ethanol series and embedded in Spurr's resin. For immunocytochemical localization, postfixation was omitted and the embedding resin used was London Resin White. Immunolabelling was carried out on ultrathin sections mounted on nickel grids and incubated overnight at 4°C with anti-FLAG antibody or, as a negative control, anti-cucumber mosaic virus polyclonal antibody (DSMZ, Braunschweig, Germany), both at 1:1000 dilution. After washing, sections were incubated for 1 h at room temperature, with 15 nm gold-labelled goat anti-rabbit serum (1:20; British BioCell, Cardiff, UK) and stained with 2% uranyl acetate and lead citrate, before being examined with a Jeol 100SX TEM (Jeol, Japan) operating at 80 KV.

Fluorescence microscopy

Leaves from Arabidopsis plants grown for 2 weeks in soil were cut lengthwise in halves and primary veins were removed. Staining was with 3,3'-dihexyloxacarbocyanine (DiOC6, Molecular Probes) at a concentration of 0.5 $\mu g\ ml^{-1}$ in PBS for 10 min, followed by washing three times in PBS. Small sections of stained leaves were placed on a microscope slide and visualized with a

63x oil immersion objective mounted on a Zeiss Axiovert 200 microscope (Carl Zeiss) equipped for epifluorescence. Simultaneous visualization of DiOC6 stain (488 nm excitation/520 nm emission) and bright-field (visible lamp) was performed using the sequential scanning facility of the microscope. Images were assembled with Adobe Photoshop software 10.0.

Results

A proportion of 16γz present in maize PB is solubilized by alcohol

16γz can be efficiently solubilized from maize endosperm by 70% ethanol supplied with reducing agent (Kim *et al.*, 2006), but its solubility in each of these agents is less clear. Treatment of purified PBs with buffer containing 2 mM DTT efficiently solubilized 27γz and 50γz, but not PB polypeptides with molecular mass in the 16 kD range (Vitale *et al.*, 1982). We therefore analyzed in more detail the solubility of 16γz accumulated in maize. An endosperm PB fraction prepared by sucrose gradient was first treated with buffer containing 4% 2-mercaptoethanol (2-ME buffer, Fig. 2A). This reducing buffer solubilizes recombinant 27γz (Mainieri *et al.*, 2014).The polypeptides with underlined fonts in Fig. 2 were identified by LC-ESI-MS/MS analysis (see Supplementary Fig. S1 and its associated Methods, and Supplementary Table S1; 27γz and 16γz identities were confirmed both in the soluble and insoluble fractions). Normal fonts indicate zeins identified solely based on their typical SDS-PAGE migration rates (notice that prolamins migrate more slowly than expected from their sequences). The 2-ME buffer very efficiently solubilized 27γz and 50γz (Fig. 2A), as expected (Vitale *et al.*, 1982). Solubilization of 16γz was instead only partial, with most of the protein remaining in the insoluble precipitate, unlike the two other γ-zeins (Fig. 2A). α-zeins, which are alcohol-soluble (Misra *et al.*, 1976), were efficiently solubilized by 70% ethanol at 25 °C (Fig. 2B). In addition, a relevant proportion of 16γz was solubilized by ethanol, whereas 50γz and 27γz remained totally or almost totally insoluble (Fig. 2B). When PBs were sequentially extracted with buffer containing non-ionic detergent, 2 mM DTT (as in Vitale *et al.*, 1982) or 4% 2-ME, the results confirmed that γ-zeins are insoluble unless reduced and indicated that DTT was not more efficient than 2-ME in solubilizing 16γz (Fig. 2C).

Minor amounts of corn legumin-1 (CL-1), an 11S storage globulin (Woo *et al.*, 2001; Yamagata *et al.*, 2003), were extracted using non-reducing buffer containing non-ionic detergent, but most of this protein was extracted in the presence of reducing agent (Fig. 2A-C and Supplementary Fig. S1). 11S storage proteins usually accumulate in protein storage vacuoles, but the presence of CL-1 in PBs, especially at late stages of endosperm maturation, has already been observed (Arcalis *et al.*, 2010; Reyes *et al.*, 2011).

The solubility of 16γz accumulated in maize is therefore intermediate between those of α-zeins and the other γ-zeins, and distinct from those of CL-1, β- and δ-zeins (these two minor zeins are not efficiently solubilized by either solvent, Fig. 2), indicating that 16γz may have specific polymerization properties. We verified this by comparing the destinies of 16γz and 27γz expressed individually in plant cells.

Recombinant 16γz and 27γz are retained intracellularly but have different solubility

The 16γz sequence was tagged at the C-terminus with the FLAG epitope. This construct (16γzf) and similarly tagged 27γz (Mainieri *et al.*, 2014; henceforth 27γzf) were first transiently expressed in tobacco protoplasts. SDS-PAGE and protein blot with anti-FLAG antibody performed around 20 h after transfection indicated that 16γzf was recovered intracellularly, with almost no sign of secretion (Fig. 3A). Besides the expected abundant monomers, a small proportion of 16γzf was detected as what appear to be dimers and larger oligomers, not disassembled by the denaturation buffer. Both the lack of secretion and the incomplete disassembly by the SDS-PAGE denaturing/reducing buffer are also a characteristic of 27γzf expressed in protoplasts (Fig. 3A, and see Mainieri *et al.*, 2014) and leaves of transgenic Arabidopsis (Geli *et al.*, 1994). Sequential extraction with non-reducing buffer and then buffer supplemented with 4% 2-ME indicated that 27γzf was almost completely insoluble unless reduced (Fig. 3A, S2 fraction), as already established (Mainieri *et al.*, 2014). A relevant proportion of 16γzf molecules was instead soluble also in the absence of reducing agent (Fig.3A, S1 fraction), indicating inefficient formation of insoluble polymers. 70% ethanol did not solubilize either of the two constructs (Fig 3B, SE fraction). When 27γzf and 16γzf were transiently co-expressed, both were almost completely insoluble in non-reducing buffer or 70% ethanol (Fig. 3B, I fraction). Therefore, the two γ-zeins interact, and 27γzf has a dominant effect in inhibiting 16γzf solubility in the absence of reducing agent. When the first buffer of the

sequential extraction was supplemented with 4% 2-ME, both individually expressed and co-expressed γ-zeins were fully solubilized, confirming the role of disulfide bonds in determining insolubility (Fig. 3C, S1 fraction). These data are consistent with the insolubility of 16γz when natural maize PBs were treated with non-reducing buffer (see Fig. 2C) and suggest that its partial solubility in ethanol is due to interactions with α-zeins. The relative proportions of monomers and oligomers detected by SDS-PAGE varied in independent experiments, but their different solubility in non-reducing conditions, when individually expressed, was consistently observed (compare 3A and 3B, and see also Supplementary Fig. S2).

In transgenic Arabidopsis, 16γzf is mostly unable to assemble into subcellular structures with the typical PB density

To compare the long-term destinies of the two zeins, the tagged constructs were expressed in transgenic Arabidopsis under constitutive promoter. These plants did not show visually evident phenotypes or clear alterations in growth and reproduction. For each construct, accumulation in leaves varied in different independent transgenic plants, but the electrophoretic pattern was unaffected by the level of final accumulation (Fig. 4A). 16γzf showed the same electrophoretic patterns if extracted from transgenic leaves or transiently transfected protoplasts, while most 27γzf monomers were clearly of higher apparent molecular mass in transgenic leaves (around 40kD), with only a minor proportion migrating as in transient expression (around 30 kD, compare Figures 3A and 4A). This indicates 27γzf-specific post-translational modifications, not yet detectable during the first hours after synthesis and not occurring in maize seeds. Hydroxylation of proline residues is the most likely explanation, as previously observed (Geli *et al.*, 1994; Mainieri *et al.*, 2014).

Subcellular localization was first investigated by isopycnic ultracentrifugation of homogenates prepared in the absence of detergent, to maintain membrane integrity. 27γzf accumulated mainly in structures with density around 1.29 (Fig. 4B). This is consistent with the known ability of 27γz to form homotypic PBs in the absence of the other zeins (Geli *et al.*, 1994; Coleman *et al.*, 1996), and the known high density of zein or zeolin PBs, in maize or transgenic plants (Larkins and Hurkman, 1978; Geli *et al.*, 1994; Mainieri *et al.*, 2004). Much lower amounts of 27γzf, probably constituted by newly-synthesized molecules not yet assembled into dense PBs, were recovered in lighter subcellular fractions that contain the ER resident endoplasmin (Klein *et al.*, 2006) and have the typical ER density (Fig. 4B). 16γzf was similarly present in the two

distinct subcellular fractions, but most of the protein was in this case in the endoplasmin-containing ER, suggesting poor ability to form PBs (Fig. 4C).

To determine the solubility of 16γzf or 27γzf present at the two positions along the gradient, fractions around 1.19 or 1.29 density were pooled, extracted with buffer without reducing agent and centrifuged to separate soluble and insoluble proteins. Around 50% of 16γzf present in the less dense fraction was solubilized by this treatment (Fig. 4D, S1), whereas nearly 100% of 16γzf or 27γzf present in fractions at 1.29 density was insoluble (Fig. 4D, P1). Treatment of P1 with 70% ethanol did not solubilize 16γzf or 27γzf (Fig. 4D, SE and PE; notice that treatment with ethanol makes the denaturation of oligomers more difficult). We conclude that the relevant proportion of 16γzf that is not assembled into dense subcellular structures is in part soluble also in the absence of reducing agent, but no 16γzf molecules insoluble in aqueous buffer are alcohol-soluble. This strongly suggests that 16γz in maize PBs is partially alcohol-soluble due to association with alcohol-soluble α-zeins, as also suggested by the data in Fig. 3B.

When homogenates, prepared in non-reducing buffer supplemented with non-ionic detergent, were subjected to velocity sucrose gradient ultracentrifugation, both 27γzf and zeolin migrated at the bottom of tubes, indicating that they are large polymers (Mainieri *et al.*, 2004, 2014). Given the partial different subcellular localization and solubility of 16γzf with respect of 27γzf, we investigated whether 16γzf also forms large polymers held together by disulfide bonds. Two plants accumulating different amounts of 16γzf were analyzed, to verify whether expression levels influence oligomerization (Fig. 5A). 16γzf migrated to the bottom of velocity ultracentrifugation tubes, independently of its level of accumulation (Fig. 5B, bottom panels). When leaf homogenization and velocity centrifugation were performed in reducing conditions, 16γzf migrated in a position corresponding to monomers (Fig. 5B, top panels). We conclude that 16γzf forms extensive, disulfide-dependent polymers, in spite of its poor ability to form high-density subcellular compartments. We therefore compared by electron microscopy the subcellular structures formed by 27γzf and 16γzf.

16γzf polymerizes into unusual reticular threads that markedly alter ER morphology

Besides typical ER membranes (Fig. 6, ER in panel A, and compare with wild-type tissue in D), 27γzf leaf tissue showed electron-dense, round-shaped structures with diameter from a few hundred nanometers to more than one micron, with attached ribosomes (Fig. 6, PB). These

structures, not present in wild-type plants, were labelled by anti-FLAG antibody (Fig. 6, panels B and C), thus indicating that 27γzf formed PBs. Homotypic PBs formed by recombinant 27γz had been observed in Arabidopsis vegetative tissues (Geli *et al.*, 1994) and tobacco seeds (Coleman *et al.*, 1996), although with sizes smaller than those we observed.

Markedly different structures were formed by 16γzf (Fig. 7). Large, irregular dilatations enclosed by a single membrane, often several micrometers wide, were detected (Fig. 7A; in Supplementary Fig. S3, black arrowheads mark the margins of this dilatation). The boundary membrane was surrounded by ribosomes (arrows in Fig. 7A, bottom enlarged inset, and Fig. 7C, enlarged inset) and connections with tubular ER were occasionally seen (white arrowheads in Supplementary Fig. S3). The vacuole was often pressing against these dilatations, sometimes leaving space for a thin layer of cytoplasm outside the dilated ER (visible in the post-fixed sample in Fig. 7A, and more easily in Supplementary Fig. S3 where the ER membrane is indicated). The lumen of ER dilatations contained very extensive electron-dense structures of two forms: very electron-opaque, osmiophilic not-oriented threads of various lengths and irregular orientation (well represented by Fig. 7A and 7C) were mainly observed, whereas a minor proportion formed more compact irregular structures of lighter electron-density (Fig. 7B, enlarged inset, and more evident in Fig. 7D and 7E). In non-osmicated tissues immunolabelled with anti-FLAG antibody, the convolutions appeared less sharp; however, gold particles were mostly aligned on them (Fig. 7, panels B and D). No labelling occurred using an irrelevant antibody, confirming that the structures are formed by 16γzf (Fig. 7E). The relative abundance of the two types of structures was variable in different ER dilatations, but when independent transgenic plants accumulating high (Fig. 7A, B, D, E) or low (Fig. 7C) amounts of 16γzf were compared, no clear relationship between recombinant protein abundance and the type of 16γzf structure could be established.

The ER vital lipophilic dye DiOC6 also efficiently stains PBs, in both rice and maize developing endosperm cells, probably due to its high affinity for the hydrophobic prolamin polypeptides (Muench *et al.*, 2000; Washida *et al.*, 2004). To complement the observations of electron microscopy, leaves were incubated with DiOC6 and observed under conventional fluorescence microscopy (Fig. 8). In 16γzf leaves, DiOC6 highlighted enlarged structures of various sizes (Fig. 8A). Higher magnification (Fig. 8A, inset, and magnification in D) showed that their content was not uniform, consistent with the structures observed by electron microscopy (Fig. 8D, arrow and compare with Fig. 7). In 27γzf leaves, more uniformly stained PBs with the classical size and round-shaped morphology were visible, as expected (Fig. 8G, and arrows in 8L). Structures similar to those in 16γzf and 27γzf were not detected in wild-type tissue, even at

very high camera exposure time that highlighted the cell periphery, as expected for the ER lipophilic dye. Both the 16γzf structures and the 27γzf PBs were also detected under transmitted light (Fig. 8E and M, arrows).

We conclude that 16γzf, unlike 27γz, is unable to form PBs and instead polymerizes into novel electron dense structures that mostly appear as irregular threads and cause marked enlargement of the ER lumen.

The N-terminal domain of 16γzf is responsible for the inefficient formation of insoluble polymers

To identify the structural features of 16γz that do not allow efficient formation of insoluble polymers, we measured the loss of solubility during pulse-chase labelling in transiently transfected tobacco protoplasts. After pulse labelling for 1h with a mixture of [^{35}S]Met and [^{35}S]Cys, protoplasts were subjected to chase for 0, 4 or 8h. At each time-point, protoplasts were directly extracted in reducing conditions (thus solubilizing all molecules of each construct, to measure synthesis and stability, Fig. 9, panels A and B), or sequentially extracted: first in non-reducing buffer and then treating the insoluble material with reducing buffer (to calculate at each time-point the percentage of molecules that are insoluble unless reduced, panel C). Each extract was immunoselected with anti-FLAG antibody and analyzed by SDS-PAGE and radiography. Newly synthesized 16γzf and 27γzf had the expected molecular mass (Fig. 9A, 16γzf and 27γzf). 16γzf was slightly more stable during the chase (Fig. 9B; data are from two fully independent experiments). Already at 0 h chase, a much higher percentage of 27γzf than 16γzf was insoluble unless reduced (Fig. 9C, 16γzf and 27γzf,). Insolubility increased during the chase, but the marked difference between the two zeins remained, as expected from the previous solubility assays (Fig. 3 and 4D). To map the insolubility determinant, we therefore prepared two constructs, 27/16 and 16/27, in which the N-terminal domain of each zein was exchanged with that of the other (Fig. 1, the green arrowhead indicates the point of exchange). Since most of the molecular mass difference between the two zeins is due to their N-terminal domain, the SDS-PAGE migrations of 27/16 and 16/27 are similar to those of 27γzf and 16γzf, respectively (Fig. 9A). The replacement of the natural N-terminal domain of 27γzf with that of 16γzf markedly inhibited insolubilization (Fig. 9C, compare 27γzf and 16/27), whereas the reciprocal replacement markedly stimulated this process (Fig. 9C, compare 16γz and 27/16). This indicates that the N-terminal domain is the major determinant for the different behavior of the two zeins.

Discussion

Mutations and insertions in the much older seed storage proteins of the 2S albumin class have been the first events originating prolamins (Pedrazzini *et al.*, 2016; Xu and Messing, 2009; Gu *et al.*, 2010). This has led to the assembly in PBs and a change in the subcellular compartment of permanent accumulation, from the vacuole to the ER, particularly in rice and panicoid cereals such as maize, sorghum and millet (Lending and Larkins, 1989; Shewry and Halford, 2002; Saito *et al.*, 2012).

16γz originated upon maize WGD (Xu and Messing, 2008) and is mainly characterized by deletions in the N-terminal region of 27γz, the most ancient γ-zein. We have shown here that recombinant 16γzf ectopically expressed in vegetative tissues accumulates within the ER, forming unusual structures. These do not resemble PBs or other ER-located polymers formed by natural or recombinant proteins expressed in plants (Bagga *et al.*, 1995; Mainieri *et al.*, 2004; de Virgilio *et al.*, 2008; Conley *et al.*, 2009; Saito *et al.*, 2009; Torrent *et al.*, 2009; Llop-Tous et al, 2010). 16γzf structures mainly consist of extensive, convoluted but well-defined filamentous threads; more rarely, the enlarged ER also contains irregular, homogenously electron-dense sectors, which may represent the proportion of 16γzf that has become insoluble. At the onset of prolamin accumulation, initial irregular dilatations along the ER were observed in rice, but with diameters below 1μM (Kawagoe *et al.*, 2005). 8S globulin, a mung bean vacuolar storage protein, formed 0.2-0.6 μM ER enlargements in transgenic tobacco BY2 cells, or, as a GFP fusion, in Arabidopsis vegetative tissues and young developing seeds, to be correctly deposited in Arabidopsis storage vacuoles only at later seed development (Wang *et al.*, 2013). The sizes of these ER structures are one order of magnitude smaller compared the dilatations caused by 16γzf. Wider, irregular ER enlargements were formed by the expression of the N-terminal region of 27γz in Arabidopsis, but these had homogeneous electron density, with no signs of filaments (Geli *et al.*, 1994). PB formed by chimeric fusions containing spider elastin-like polypeptide can have a loosely packed content, but they are round-shaped and rarely larger than 3 μM, with no well-defined filaments (Conley *et al.*, 2009; Phan *et al.*, 2014). Therefore, the unusual structures formed by 16γzf markedly differ from 27γzf PBs and from ER enlargements formed by certain storage proteins at early stages of seed development or by protein fusions that polymerize in the ER.

However, 16γzf threads strikingly resemble those formed by diabetes insipidus-inducing mutants of the antidiuretic hormone arginine vasopressin precursor (Birk *et al.*, 2009; Beuret *et al.*, 2017). These dominant mutations can be in different locations along the precursor, but they all result in abnormal inter-chain disulfide-bonds leading to oligomerization and in some cases partial resistance to denaturation by SDS/reducing agent, whereas the normal precursor has eight intra-chain bonds. The misfolded precursors thus accumulate in the ER instead of trafficking to secretory granules and form irregularly packed electron-dense filaments, which in some cases coalesce in more uniformly electron-dense regions, similarly to 16γzf. Although the mutated precursors seem unable to form canonical amyloid cross-β-sheets, their ability to form fibers resembles amyloid aggregation (Beuret *et al.*, 2017).

Our data indicate that 16γzf is not a structurally defective protein rapidly degraded by ER quality control. 16γzf threads are disulfide-bonded polymers that remain partially soluble in oxidizing conditions, unlike 27γzf polymers. Only upon co-expression of the two recombinant zeins 16γzf becomes fully insoluble unless reduced, indicating direct interactions with 27γzf. 16γz present in natural maize PBs is in part solubilized by alcohol together with α-zeins, but no alcohol-soluble 16γzf is detected in transgenic Arabidopsis or upon transient co-expression with 27γzf. This supports the hypothesis that, in maize, at least one of the alcohol-soluble α-zeins directly interacts with16γz, consistent with the location of 16γz in natural PBs (Lending and Larkins, 1989; Yao *et al.*, 2016) and the results of two-hybrid yeast assays (Kim *et al.*, 2002, 2006). A specific role of 16γz in natural, heterotypic PB assembly is also supported by the characteristics of two maize mutations with opaque endosperm, *mucronate* and *opaque10*. *Mucronate* is a frameshift mutation that completely changes the 16γz sequence for the last 63 amino acids, abolishes its solubility in 70% ethanol supplemented with 2-ME, and markedly weakens the interaction with 22kD α-zein (Kim *et al.*, 2006). In *mucronate* seeds, the overall amount of zeins is reduced (Salamini *et al.*, 1983), and PBs have angular deformations that often interrupt the outer layer, indicating defects in the organization of the interface between α– and γ-zeins (Zhang and Boston, 1992). *Opaque10* is a frameshift mutation generating a premature stop codon in a cereal-specific protein located in PBs (Yao *et al.*, 2016). *Opaque10* PBs are misshaped and often irregularly elongated. The ordered localizations of 16γz and of the 22kD α-zein that normally is located next to it are disrupted, and the two zeins are dispersed in the PB (Yao *et al.*, 2016). RNA interference, used to inhibit the synthesis of γ-zeins in maize, also caused PB misshaping and angular deformations (Wu and Messing, 2010). A specific role of 16γz could not be established in this case, since the synthesis of both the 27 and 16kD polypeptides was almost fully inhibited. However, RNA interference in which the synthesis of

16γz, 50γz and β-zein was concomitantly suppressed indicated that these proteins are mainly involved in PB expansion, whereas 27γ-zein controls PB initiation and shape, consistent with our data in transgenic Arabidopsis (Guo *et al.*, 2013).

Sorghum (*Sorghum bicolor*), a very close relative of maize (Swigoňová *et al.*, 2004), has not undergone WGD and contains only two genes belonging to the same prolamin II group of γ-zeins, kafirin2γ27 and kafirin2γ50 (Belton *et al.*, 2006; Xu and Messing, 2009), therefore lacking a 16γz orthologue. Similarly to β– and γ-zeins, β– and γ-kafirins form the more electron-dark structures of the PB, which however are not limited to the PB periphery and are also concentrated in the central core or form patches within the lighter regions (Shull *et al.*, 1992). This less ordered contact of darker and lighter regions (the latter mainly containing α-type prolamins) compared to maize PBs may thus be related to the absence of a 16γz-like prolamin.

Our domain exchange results suggest that the different behavior of the two γ-zeins is mainly due to their N-terminal domains. A synthetic version of the $(VHLPPP)_8$ repeated segment has an amphipathic polyproline II structure and in vitro affinity to liposomes that partially mimic the lipid composition of the plant ER, suggesting that the repeat may favor interaction of 27γz with the inner surface of the ER membrane (Kogan *et al.*, 2004). The Zera sequence is a 27γz portion almost identical to the one used to construct zeolin and, like zeolin, it determines PB formation in a Cys-dependent fashion when fused to a number of proteins (Torrent *et al.*, 2009). In a Zera-fluorescent protein fusion, progressive deletion of the Pro-rich hexapeptides led to progressively increased secretion and reduced PB size but did not alter their spherical shape (Llop-Tous *et al.*, 2010), indicating that the peculiar structures formed by 16γz are not simply due to the loss of repeats. Indeed, the N-terminal region of 16γz has also lost three Cys residues and contains two degenerated Pro-rich sequences containing two new Tyr residues - aromatic amino acids inhibit the formation of polyproline II helices (Brown and Zondlo, 2012) - as well as other aromatic amino acids and a new Gln-rich short sequence (Fig. 1). Altogether, these features may have abolished the ability to orderly interact with lipids and determined the formation of rod-like polymers involved in stabilizing the γ-zein/α-zein interface.

Proteins containing disulfide bonds have generally higher evolutionary rates (Feyertag and Alvarez-Ponce, 2017). Intra-chain disulfides probably stabilize important conformations and thus have a buffering, chaperone-like, effect making the polypeptide more tolerant to mutations; thus, once acquired, inter-chain disulfides are rarely changed (Feyertag and Alvarez-Ponce, 2017; Wong *et al.*, 2011). Unpaired Cys residues are also relatively more conserved than other amino acids (Wong *et al.*, 2011). The major deletion and the mutations generating 16γz have

eliminated a number of 27γz cysteine residues and have altered the biochemical and polymerization properties of the prolamin, but they have not caused gross misfolding and degradation by quality control. They have instead promoted a new role of the protein and a new PB organization.

Prolamins form peculiar heteropolymers. Analysis of many prolamin polypeptides and their positioning within a PB in different grasses indicates that a high genetic variability is tolerated, probably because PB function is simply constituted by the high accumulation of reduced nitrogen in the first compartment of the secretory pathway. However, within an individual species, certain requirements for optimal PB assembly exist, as indicated by the many natural and artificial cereal mutants analyzed. We have shown here that an apparently defective zein polypeptide, generated upon maize WGD, forms very unusual structures that may explain its specific structural role at the interface between the ancient and the more recently evolved maize prolamins. The organization of 16γz structures resemble abnormally disulfide-linked, amyloid-like fibers formed by pathological mutants of a human hormone precursor. It thus appears that mutations giving rise to similar abnormal structures within the ER can result in pathogenic loss of function in one case but exploited in a developmental process in another one.

Acknowledgments

This work was supported by Project “Risorse biologiche e tecnologie innovative per lo sviluppo sostenibile del sistema agroalimentare” and Project “Filagro - Strategie innovative e sostenibili per la filiera agroalimentare”, of the 2006 and 2012 Agreements between Regione Lombardia and CNR (to E.P. and A.V.).

Author Contributions

D. Mainieri, E.P., E.S., and A.V. designed research; D. Mainieri, C.A.M., B.P., D. Maffi, M.T., F.F., E.P. and A.V. performed research; E.S., F.F., E.P. and A.V. wrote the manuscript.

References

Arcalis E, Stadlmann J, Marcel S, Drakakaki G, Winter V, Rodriguez J, Fischer R, Altmann F, Stoger E. 2010. The changing fate of a secretory glycoprotein in developing maize endosperm. Plant Physiology **153,** 693–702.

Bagga S, Adams H, Kemp JD, Sengupta-Gopalan C. 1995. Accumulation of 15-kilodalton zein in noveL protein bodies in transgenic tobacco. Plant Physiology **107,**13-23.

Belton PS, Delgadillo I, Halford NG, Shewry PR. 2006. Kafirin structure and functionality. Journal of Cereal Science **44,** 272-286.

Beuret N, Hasler F, Prescianotto-Baschong C, Birk J, Rutishauser J, Spiess M. 2017. Amyloid-like aggregation of provasopressin in diabetes insipidus and secretory granule sorting. BMC Biology **15,** 5.

Birk J, Friberg MA, Prescianotto-Baschong C, Spiess M, Rutishauser J. 2009. Dominant pro-vasopressin mutants that cause diabetes insipidus form disulfide-linked fibrillar aggregates in the endoplasmic reticulum. Journal of Cell Science **122,** 3994–4002.

Brown AM, Zondlo NJ. 2012. A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions. Biochemistry **51,** 5041-5051.

Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proceedings of the National Academy of Sciences, USA **103,** 2730–2735.

Coleman CE, Herman EM, Takasaki K, Larkins BA. 1996. The maize γ-zein sequesters α-zein and stabilizes its accumulation in protein bodies of transgenic tobacco endosperm. The Plant Cell **8,** 2335-2345.

Conley AJ, Joensuu JJ, Menassa R. 2009. Induction of protein body formation in plant leaves by elastin-like polypeptide fusions. BMC Biology 7, 48.

Conley AJ, Joensuu JJ, Menassa R, Brandle JE. 2009. Induction of protein body formation in plant leaves by elastin-like polypeptide fusions. BMC Biology **7,** 48.

de Virgilio M, De Marchis F, Bellucci M, Mainieri D, Rossi M, Benvenuto E, Arcioni,S, Vitale A. 2008. The human immunodeficiency virus antigen Nef forms protein bodies in leaves of transgenic tobacco when fused to zeolin. Journal of Experimental Botany **59,** 2815-2829.

Ems-McClung SC, Benmoussa M, Hainline BE. 2002. Mutational analysis of the maize gamma zein C-terminal cysteine residues. Plant Science **162,** 131–141.

Faoro F, Tornaghi R, Belli G. 1991. Localization of the Closteroviruses on grapevine thin-section and their identification by immunoglod labelling. Journal of Phytopathology-Phytopathologische Zeitschrift **133,** 297-306.

Feyertag F, Alvarez-Ponce D. 2017. Disulfide bonds enable accelerated protein evolution. Molecular Biology and Evolution **34,** 1833–1837.

Geli MI, Torrent M, Ludevid D. 1994. Two structural domains mediate two sequential events in γ-zein targeting: protein endoplasmic reticulum retention and protein body formation. The Plant Cell **6,** 1911–1922.

Gomez-Navarro N, Miller E. 2016. Protein sorting at the ER-Golgi interface. Journal of Cell Biology **215,** 769-778.

Gu YQ, Wanjugi H, Coleman-Derr D, Kong X, Anderson OA. 2010. Conserved globulin gene across eight grass genomes identify fundamental units of the loci encoding seed storage proteins. Functional & Integrative Genomics **10,** 111–122.

Guo X, Yuan L, Chen H, Sato SJ, Clemente TE, Holding DR. 2013. Nonredundant function of zeins and their correct stoichiometric ratio drive protein body formation in maize endosperm. Plant Physiology **162,** 1359–1369.

Jiao YN, Wickett NJ, Ayyampalayam S, *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. Nature **473,** 97-100.

Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA. 2009. Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. Genome Research **19,** 1404–1418.

Kawagoe Y, Suzuki K, Tasaki M, *et al.* 2005. The critical role of disulfide bond formation in protein sorting in the endosperm of rice. The Plant Cell **17,** 1141–1153.

Kim CS, Gibbon BC, Gillikin, JW, Larkins BA, Boston RS, Jung R. 2006. The maize *Mucronate* mutation is a deletion in the 16-kDa gamma-zein gene that induces the unfolded protein response. Plant Journal **48,** 440-451.

Kim CS, Woo Ym YM, Clore AM, Burnett RJ, Carneiro NP, Larkins BA. 2002. Zein protein interactions, rather than the asymmetric distribution of zein mRNAs on endoplasmic reticulum membranes, influence protein body formation in maize endosperm. The Plant Cell **14,** 655–672.

Klein E M, Mascheroni L, Pompa A, Ragni L, Weimar T, Lilley KS, Dupree P, Vitale A. 2006. Plant endoplasmin supports the protein secretory pathway and has a role in proliferating tissues. Plant Journal **48,** 657–673.

Kogan, M. J., Lopez, O., Cocera, M., Lopez-Iglesias, C., De La Maza, A., Giralt, E. 2004. Exploring the interaction of the surfactant N-terminal domain of gamma-Zein with soybean phosphatidylcholine liposomes. Biopolymers **73,** 258–268.

Larkins BA, Hurkman WJ. 1978. Synthesis and Deposition of Zein in Protein Bodies of Maize Endosperm. Plant Physiology **62,** 256-263.

Lending CR, Larkins BA. 1989. Changes in the zein composition of protein bodies during maize endosperm development. The Plant Cell **1,** 1011–1023.

Llop-Tous I, Madurga S, Giralt E, Marzabal P, Torrent M, Ludevid MD. 2010. Relevant elements of a maize γ-zein domain involved in protein body biogenesis. Journal of Biological Chemistry **285,** 35633–35644.

Mainieri D, Morandini F, Maîtrejean M, Saccani A, Pedrazzini, Vitale A. 2014. Protein body formation in the endoplasmic reticulum as an evolution of storage protein sorting to vacuoles: insights from maize γ-zein. Frontiers in Plant Science **5,** 331.

Mainieri D, Rossi M, Archinti M, Bellucci M, De Marchis F, Vavassori S, Pompa A, Arcioni S, Vitale A. 2004. Zeolin: a new recombinant storage protein constructed using maize γ-zein and bean phaseolin. Plant Physiology **136,** 3447–3456.

Misra PS, Mertz ET, Glover DV. 1976. Studies on corn proteins. IX. Comparison of the amino acid composition of Landry-Moureaux and Paulis-Wall endosperm fractions. Cereal Chemistry **53,** 699-704.

Muench DG, Chuong SDX, Franceschi VR, Okita TW. 2000. Developing prolamine protein bodies are associated with the cortical cytoskeleton in rice endosperm cells. Planta **211,** 227-238.

Pedrazzini E, Mainieri D, Marrano CA, Vitale A. 2016. Where do protein bodies of cereal seeds come from? Frontiers in Plant Science **7,** 1139.

Phan HT, Hause P, Hause G, Arcalis E, Stoger E, Maresch D, Altmann F, Joensuu J, Conrad U. 2014. Influence of elastin-like polypeptide and hydrophobin on recombinant hemagglutinin accumulations in transgenic tobacco plants. PlosONE **9,** e99347.

Pompa A, Vitale A. 2006. Retention of a bean phaseolin/maize γ-zein fusion in the endoplasmic reticulum depends on disulfide bond formation. The Plant Cell **18,** 2608–2621.

Prat S, Cortadas J, Puigdomènech P, Palau J. 1985. Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2. Nucleic Acids Research **13,** 1493-1504.

Prat S, Pérez-Grau L, Puigdomènech P. 1987. Multiple variability in the sequence of a family of maize endosperm proteins. Gene **52,** 41-49.

Reyes FC, Chung T, Holding D, Jung R, Vierstra R, Otegui M. 2011. Delivery of prolamins to protein storage vacuole in maize aleurone cells. The Plant Cell **23,** 769-784.

Saito Y, Kishida K, Takata K, Takahashi H, Shimada T, Tanaka K, Morita S, Satoh S, Masumura T. 2009. A green fluorescent protein fused to rice prolamin forms protein body-like structures in transgenic rice. Journal of Experimental Botany **60,** 615–627.

Saito Y, Shigemitsu T, Yamasaki R, Sasou A, Goto F, Kishida K, Kuroda M, Tanaka K, Morita S, Satoh S, Masumura T. 2012. Formation mechanism of the internal structure of type I protein bodies in rice endosperm: relationship between the localization of prolamin species and the expression of individual genes. Plant Journal **70,** 1043–1055.

Salamini F, Di Fonzo N, Fornasari E, Gentinetta E. 1983. Mucronate, Mc, a dominant gene of maize which interacts with *opaque-2* to suppress zein synthesis. Theoretical and Applied Genetics **65,** 123-128.

Shewry PR, Halford NG. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. Journal of Experimental Botany **53,** 947–958.

Shull JM, Watterson JJ, Kirleis AW. 1992. Purification and immunocytochemical localization of kafirins in *Sorghum bicolor* (L. Moench) endosperm. Protoplasma **171,** 64-74.

Swigoňová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. 2004. Close split of sorghum and maize genome progenitors. Genome Research **14,** 1916–1923.

Torrent M, Llompart B, Lasserre-Ramassamy S, Llop-Tous I, Bastida M, Marzabal P, Westerholm-Parvinen A, Saloheimo M, Heifetz PB, Ludevid MD. 2009. Eukaryotic protein production in designed storage organelles. BMC Biology **7,** 5.

Vitale A, Smaniotto E, Longhi R, Galante E. 1982. Reduced soluble proteins associated with maize endosperm protein bodies. Journal of Experimental Botany **33,** 439-448.

Wang J, Shen J, Cai Y, Robinson DG, Jiang L. 2013. Successful transport to the vacuole of heterologously expressed mung bean 8S globulin occurs in seed but not in vegetative tissues. Journal of Experimental Botany **64,** 1587–1601.

Washida H, Sugino A, Messing J, Esen A, Okita, TW. 2004. Asymmetric localization of seed storage protein RNAs to distinct subdomains of the endoplasmic reticulum in developing maize endosperm Cells. Plant & Cell Physiology **45,** 1830-1837.

Wong JW, Ho SY, Hogg PJ. 2011. Disulfide bond acquisition through eukaryotic protein evolution. Molecular Biology and Evolution **28,** 327–334.

Woo Y-M, Hu DW-N, Larkins BA, Jung R. 2001. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. The Plant Cell **13,** 2297–2317.

Wu Y, Messing J. 2010. RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. Plant Physiology **153,** 337–347.

Xu J-H, Messing J. 2008. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proceedings of the National Academy of Sciences, USA **105,** 14330–14335.

Xu J-H, Messing J. 2009. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. Theoretical and Applied Genetics **119,** 1397–1412.

Yamagata T, Kato H, Kuroda S, Abe S, Davies E. 2003. Uncleaved legumin in developing maize endosperm: Identification, accumulation and putative subcellular localization. Journal of Experimental Botany **54,** 913–922.

Yao D, Qi W, Li X, Yang Q, Yan S, Ling H, Wang G, Wang G, Song R. 2016. Maize opaque10 encodes a cereal-specific protein that is essential for the proper distribution of zeins in endosperm protein bodies. PLoS Genetics **12,** e1006270.

Yu M, Lau TY, Carr SA, Krieger M. 2012. Contributions of a disulfide bond and a reduced cysteine side chain to the intrinsic activity of the HDL receptor SR-BI. Biochemistry **51,** 10044–10055.

Zhang F, Boston RS. 1992. Increases in binding protein (BiP) accompany changes in protein body morphology in three high-lysine mutants of maize. Protoplasma **171,** 142-152.

Figure Legends

Fig. 1. Schematic cartoon and amino acid sequence of the 27kD and 16kD γ-zein primary translation products. The question mark indicates that the linkage status of the two Cys residues towards the C-terminus of 16kD γ-zein is not known. The green arrowhead indicates the point at

which the N- and C-terminal regions of the two polypeptides were exchanged, to produce constructs 27/16 and 16/27. In the amino acid sequences, Cys residues are highlighted in red.

Fig. 2. The solubility of 16γz accumulated in maize PB is intermediate between those of α-zeins and the other γ-zeins. PBs purified from maize seeds, collected at 25 DAP, were treated at 4°C with buffer containing 4% 2-ME (**A**) or at 25°C with 70% ethanol in H_2O (**B**). After centrifugation, soluble and insoluble proteins were analyzed by SDS-PAGE and Coomassie staining. The different zein polypeptides (αz, βz, γz, δz) and CL-1 are indicated. Those whose identities were confirmed by LC-ESI-MS/MS are in underlined fonts (see also Supplementary Fig. S1 and Supplementary Table S1). **C**. Purified PBs were sequentially extracted with buffer containing 1% Triton X-100 (Triton), 2mM DTT (DTT) and 4% 2-ME (2-ME). After each step, the suspension was centrifuged and the soluble material was analyzed by SDS-PAGE and Coomassie staining, together with the insoluble material of the last extraction (insol.). The positions of molecular mass markers (directly visible in the last lane of panel C, M) are indicated at right in panels B and C, in kD.

Fig. 3. Recombinant 16γz and 27γz are retained intracellularly but have different solubility. Protoplasts were isolated from tobacco leaves and transiently transformed with plasmids encoding the indicated constructs or with empty vector (Co) and analyzed after incubation for 20 h. **A.** Protoplasts (in) or incubation medium (out) were homogenized in the absence (-) of 2-ME. After centrifugation, soluble (S1) and insoluble fractions were collected. The insoluble material was resuspended in the presence (+) of 2-ME and subjected to a second centrifugation, to obtain the new soluble (S2) and insoluble (I) fractions. **B.** Protoplasts (in) or incubation medium (out) were homogenized in the absence of 2-ME. After centrifugation, soluble (S1) and insoluble fractions were collected. The insoluble material was resuspended with 70% ethanol and subjected to a second centrifugation, to obtain the new soluble (SE) and insoluble (I) fractions. **C.** As in **B**, but the first homogenization was performed in the presence of 4% 2-ME. In all panels, upper images show analysis of each fraction by SDS-PAGE and protein blot with anti-FLAG antibody, whereas lower images are Ponceau S staining. The positions of molecular mass markers are shown at left, in kD. In **B** and **C**, the positions of dimers (dim) and monomers (mon) of 27γzf (27) and 16γzf (16) are indicated.

Fig. 4. Assembly of 16γzf into dense subcellular structures is inefficient. **A.** Leaves from transgenic Arabidopsis expressing 27γzf or 16γzf, or from wild type plants (WT) were homogenized in the presence of 2-ME. Soluble proteins were analyzed by SDS-PAGE. Each individual lane represents an independent transgenic plant. Upper images are protein blots with anti-FLAG antibody; lower images are Ponceau S staining. **B** and **C**. Leaves from transgenic Arabidopsis expressing 27γzf (**A**) or 16γzf (**B**) were homogenized in the presence of 12% (w/w) sucrose and absence of detergent. The homogenates were fractionated by ultracentrifugation on 16-65% (w/w) isopycnic sucrose gradients. Proteins in each gradient fraction were analyzed by SDS-PAGE and protein blot, with anti-FLAG (27γzf, 16γzf) antibody or anti-endoplasmin (endopl.) serum. Top of gradients is at left. In each panel, numbers at top indicate density (g ml^{-1}). **D.** Fractions around either 1.19 or 1.29 density from the gradients shown in panels **B** and **C** were pooled, extracted with buffer without reducing agent and centrifuged. Supernatant (S1) and pellet (P1) were collected. An aliquot of P1 was further treated with 70% ethanol and centrifuged to obtain ethanol-soluble (SE) and insoluble (PE) material. Upper image shows analysis by SDS-PAGE and protein blot with anti-FLAG antibody; lower image is Ponceau S staining. In all panels, numbers at left indicate the positions of molecular mass markers, in kD.

Fig. 5. 16γzf forms large, disulfide-dependent polymers. **A.** Homogenates were prepared from leaves of two independent transgenic Arabidopsis lines that accumulate different amounts of 16γzf (#11 and #4, two plants for each line), and from leaves of untransformed Arabidopsis (WT), and analyzed by SDS-PAGE. The upper image is the protein blot with anti-FLAG antibody; the lower image is Ponceau S staining. **B.** Homogenates were prepared in either oxidizing or reducing buffer, and fractionated by velocity gradient ultracentrifugation. Top of each gradient is at left. T: unfractionated total homogenate, P: pellet at the bottom of the tube after centrifugation. Numbers at top indicate the positions where molecular mass markers migrate along the gradients. In all panels, numbers at left indicate the positions of SDS-PAGE molecular mass markers, in kD.

Fig. 6. 27γzf forms PBs. Leaves from six weeks old transgenic Arabidopsis plants expressing 27γzf (**A**-**C**), or wild-type plants (**D**), were analyzed by electron microscopy. **A, D.** Osmium-postfixed ultrathin section. **B, C.** Immunolabelling with anti-FLAG antibody and secondary 15 nm gold-conjugated goat anti-rabbit serum. ER, endoplasmic reticulum; PB, protein body; M, mitochondria; N, nucleus; arrows, ribosomes. Notice that in non-osmicated immunolabelled

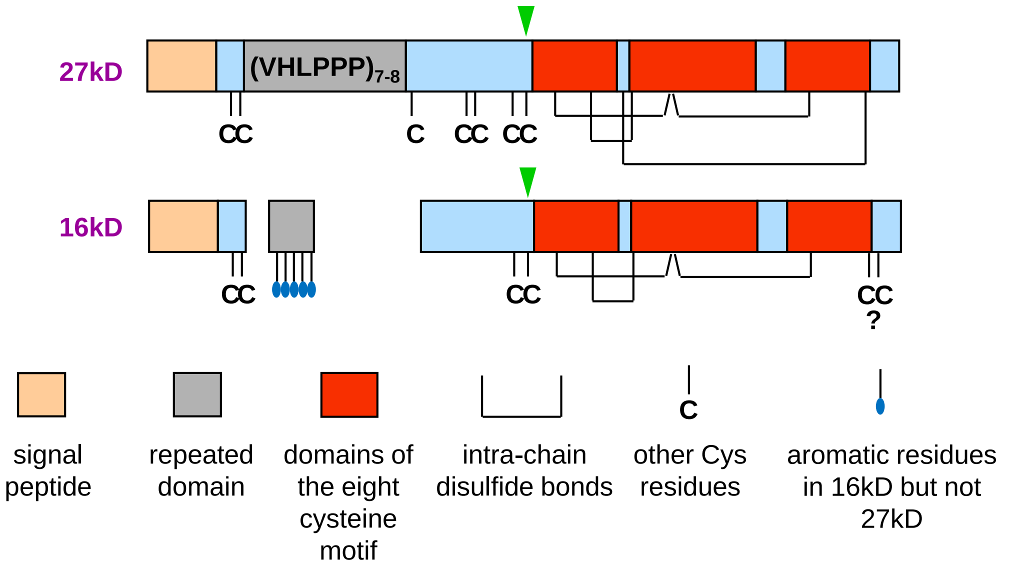
tissues (**B, C**) ER membranes are not detectable, however numerous ribosomes are visible aligned outside the PB periphery (arrows). Bars = 200 nm.

Fig. 7. 16γzf does not form PBs but forms electron-dense structures in highly enlarged endoplasmic reticulum (ER) lumen. Leaves from six weeks old transgenic Arabidopsis plants accumulating 16γzf in high (**A, B, D, E**) or low (**C**) amounts were analyzed. **A, C.** Osmium-postfixed ultrathin sections. **B, D, E** Immunolabelling with anti-FLAG antibody (**B, D**) or irrelevant antibody as negative control (**E**), and secondary 15 nm gold-conjugated goat anti-rabbit serum. Insets in panels **A, B** and **C** show magnifications, to better appreciate the ribosomes attached on the cytosolic side of the ER membrane (arrows) and the electron dense convoluted structures within the ER lumen. Ch, chloroplasts; M, mitochondria; V, vacuole. Bars in **A, D** and all insets = 200 nm; bars in **B, C, E** = 500 nm.

Fig. 8. 27γzf PBs and 16γzf ER enlargements can be detected by fluorescence staining of the ER. Leaf tissue from 16γzf (**A-F**), 27γzf (**G-N**) or wild type (**O-Q**) Arabidopsis plants was stained with the DiOC6 dye and analysed by epifluorescence microscopy. (**A,D,G,L,O**): DiOC6 fluorescence (green); (**B,E,H,M,P**): bright-field; (**C,F,I,N,Q**): merge. Camera exposure time (ms): 61 (**A**), 502 (**G**), 8352 (**O**). Boxes in **A-C** and **G-I** indicate the regions that are shown at higher magnification in **D-F** and **L-N**, respectively. The white arrows indicate enlarged ER (**D-F**) or PBs (**L-N**).

Fig. 9. The N-terminal domain of 16γzf has a major role in inhibiting the formation of insoluble polymers. Protoplasts prepared from tobacco leaves were transiently transfected with plasmids encoding the indicated constructs or with empty plasmid (Co). **A.** Transfected protoplasts were pulse labelled with radioactive amino acids for 1h before homogenization in reducing conditions. Proteins were immunoselected using anti-FLAG antibody and analyzed by SDS-PAGE and radiography. The different lanes are from a single exposure of a single radiography from which irrelevant lanes have been removed. The newly synthesized recombinant polypeptides (arrowheads) and the positions of molecular mass markers (numbers at left, kD) are indicated. **B.** Protoplasts pulse-labelled as in (**A**) were subjected to chase for the indicated h, homogenized in the presence of 2-ME, immunoselected with anti-FLAG antibody and analyzed by SDS-PAGE and radiography. For each chase time-point, density of the relevant radioactive bands was

quantified and expressed as percentage of the intensity at 0h chase. **C.** Protoplasts pulse-labelled as in (**A**) and chased for the indicated h were then subjected to sequential homogenization steps, first in the absence 2-ME and then treating with 2-ME the insoluble material. Proteins of each step were immunoselected with anti-FLAG antibody and analyzed by SDS-PAGE and radiography. At each time point, density of the relevant radioactive bands was quantified and, for each construct, expressed as percentage in the second immunoprecipitation step over the sum of the two immunoprecipitations (% insoluble). In (**B**) and (**C**), values from two fully independent transient expression experiments are shown.



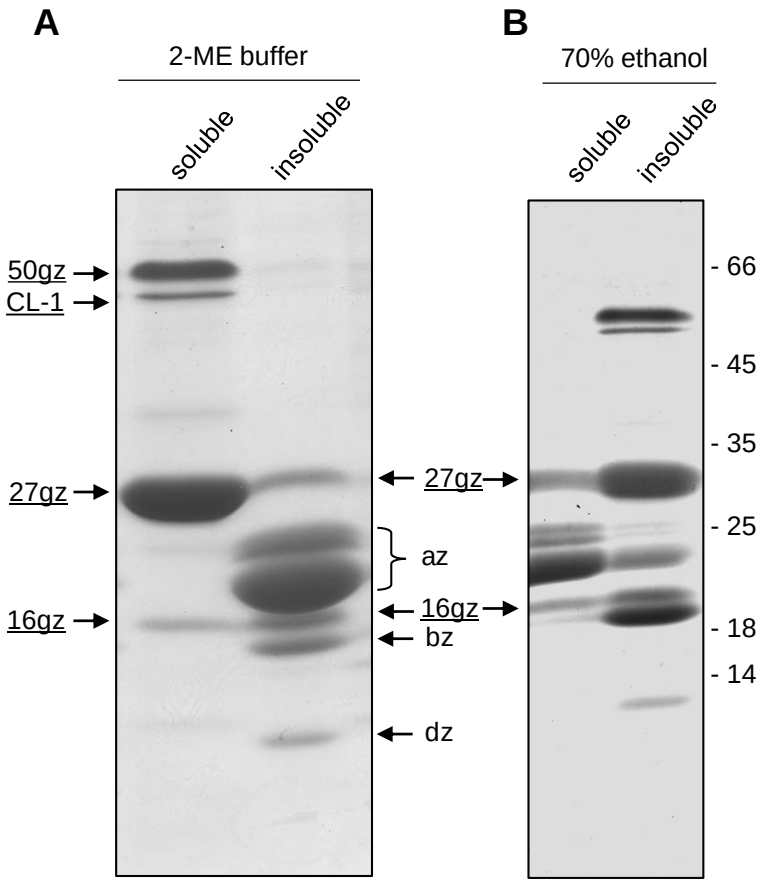

```
                    70        80        90       100       110       120
27kD  MRVLLVALALLALAASATSTHTSGGCGCQPPPPVHLPPPVHLPPPVHLPPPVHLPPPVHL
      :.::.::::::::::::::.:. :::::::::.::        ::::: ..::: .:::
16kD  MKVLIVALALLALAASAASS-TSGGCGCQTPP-------FHLPPPFYMPPPFYLPP----
               10        20         30                40

               70        80        90       100       110       120
27kD  PPPVHLPPPVHVPPPVHLPPPPCHYPTQPPRPQPHPQPHPCPCQQPHPSPCQLQGTCGVG
                                      :..::: ..: :.:. :::: :.::::
16kD  --------------------------------QQQPQPWQYPTQPPQLSPCQQFGSCGVG
                                         50        60        70

                130       140       150       160       170
27kD  S--TPILGQCVEFLRHQCSPTATPYCSPQCQSLRQQCCQQLRQVEPQHRYQAIFGLVLQS
      :  .:.:::::::::::::::.:::: :::::.:.::::.:.::::: :::::..:.::::
16kD  SVGSPFLGQCVEFLRHQCSPAATPYGSPQCQALQQQCCHQIRQVEPLHRYQATYGVVLQS
          80        90       100       110       120       130

      180       190       200         210        220
27kD  ILQQQPQSGQVAGLLAAQIAQQLTAMCGLQ--QPTPCPY-AAAGGVPH
      .::::: :..:.:.:::.:::::::::::  ::.:::. :::::: .
16kD  FLQQQPQ-GELAALMAAQVAQQLTAMCGLQLQQPGPCPCNAAAGGVYY
        140        150       160       170       180
```

Fig. 1



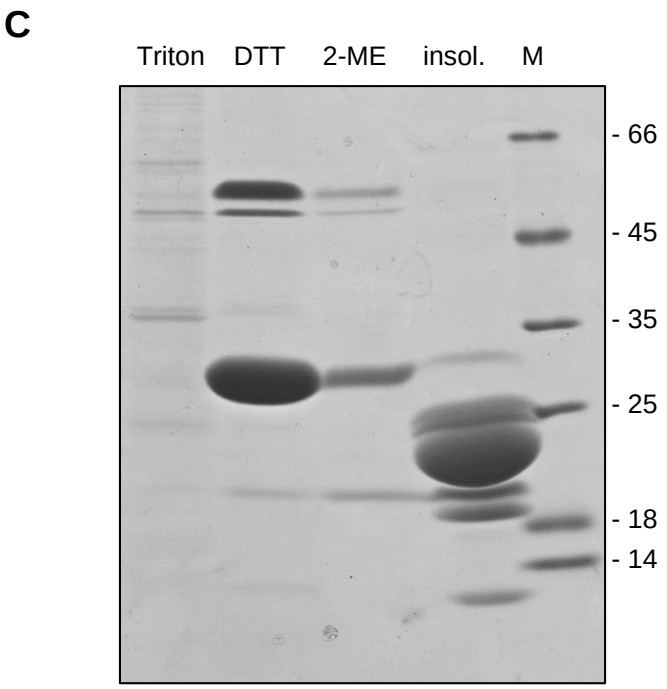


Fig. 2

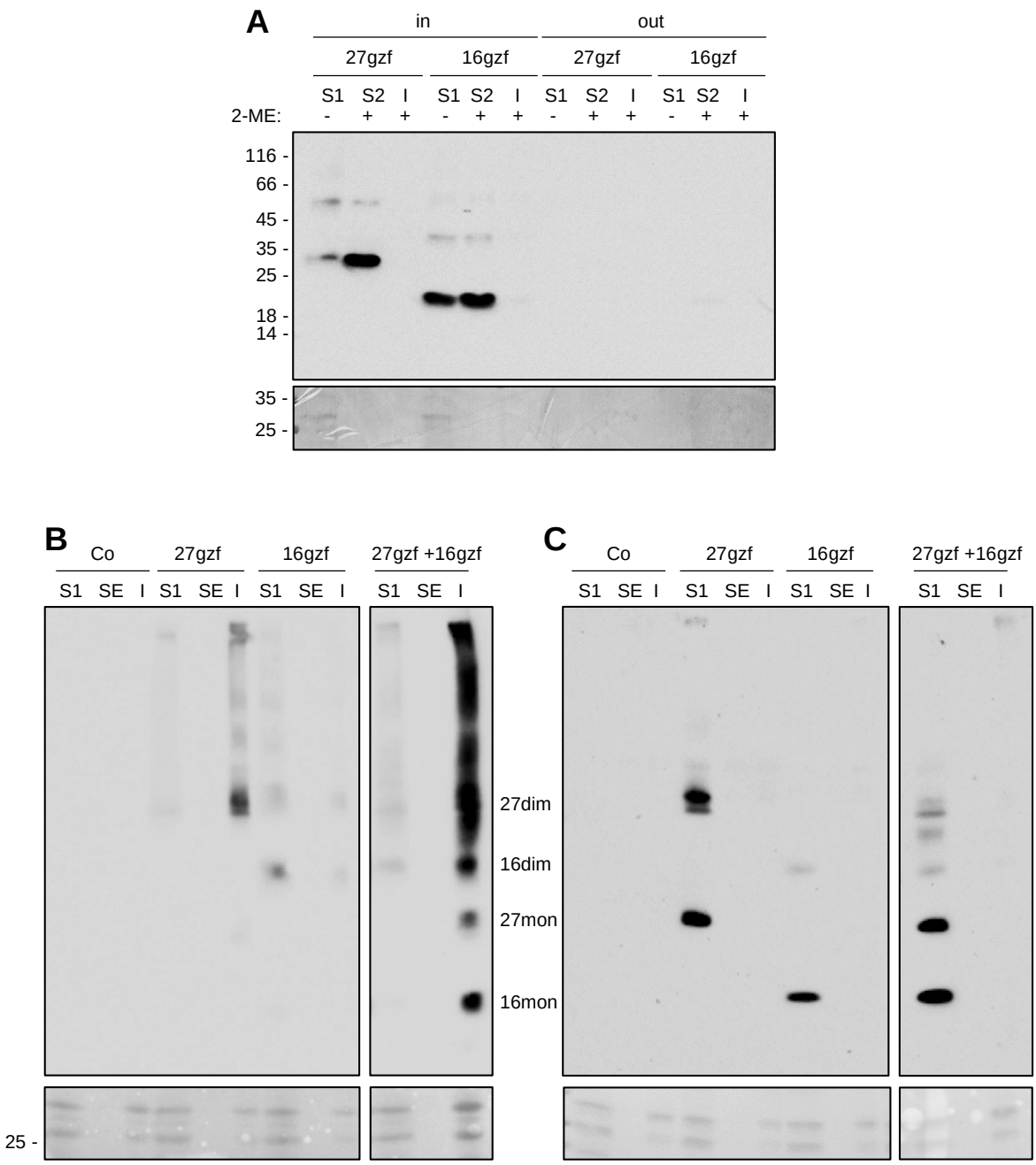


Fig. 3

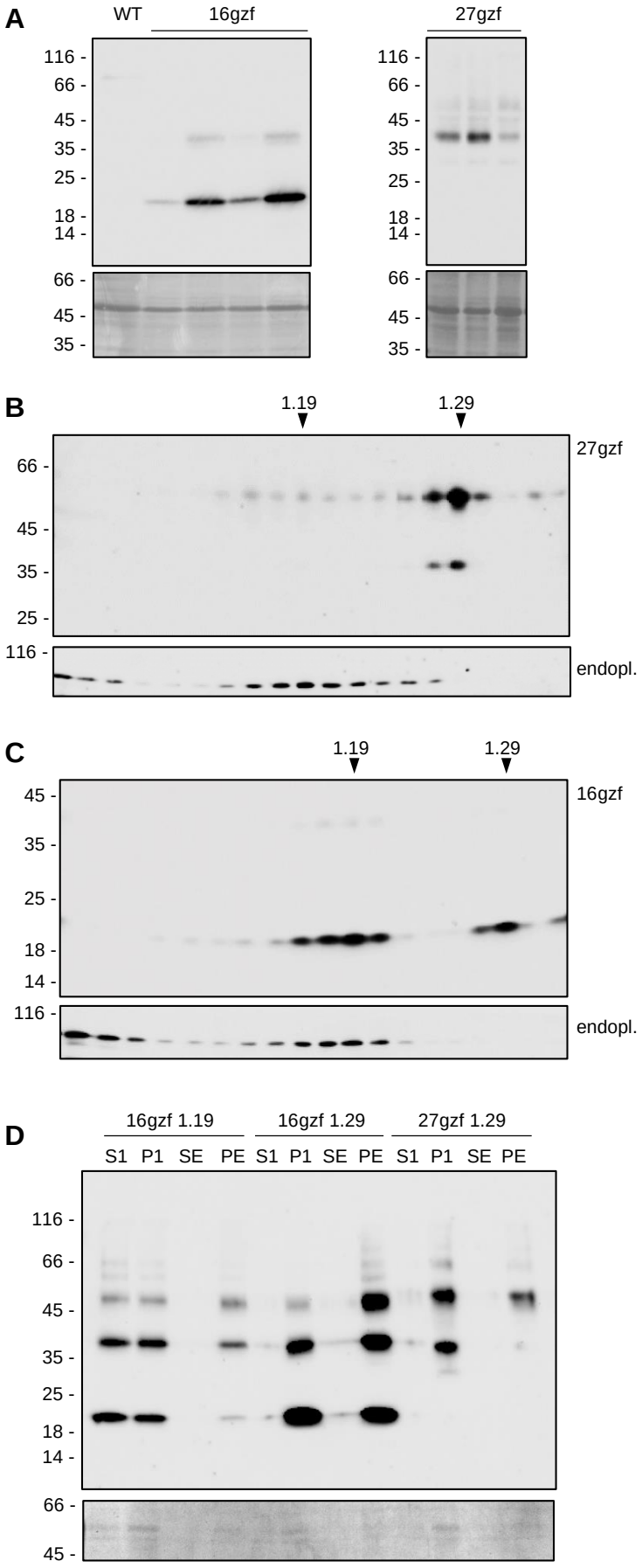


Fig. 4

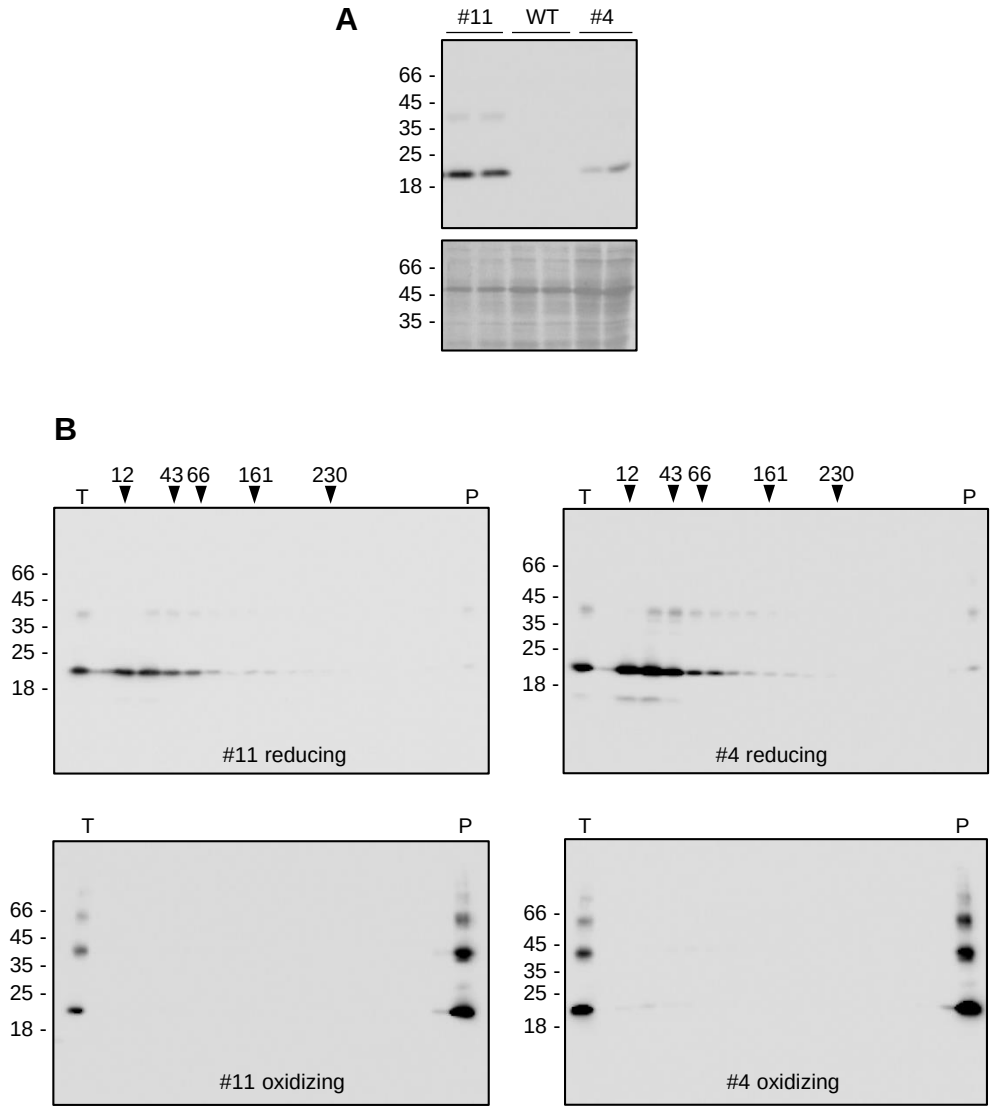


Fig. 5

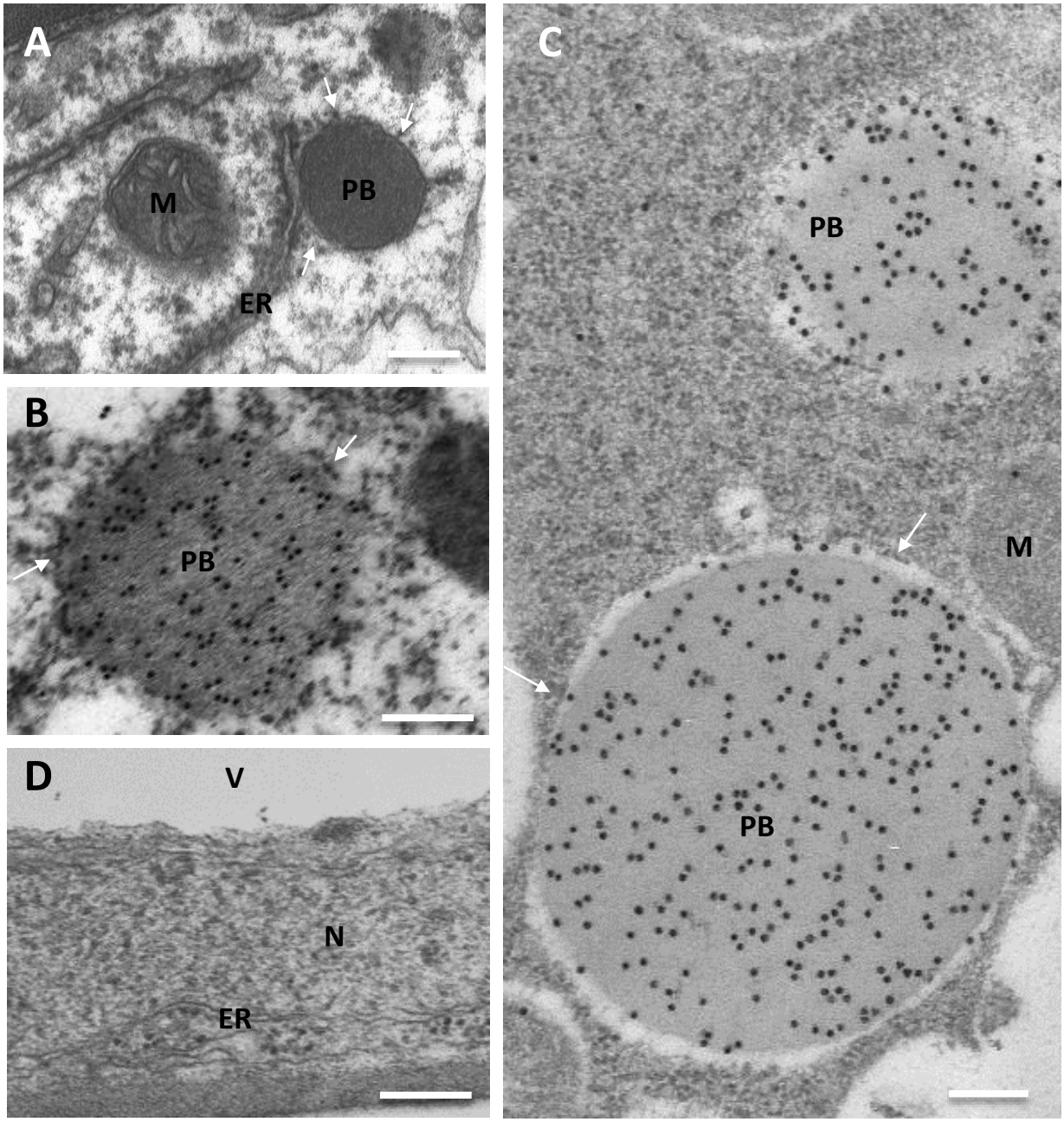


Fig. 6

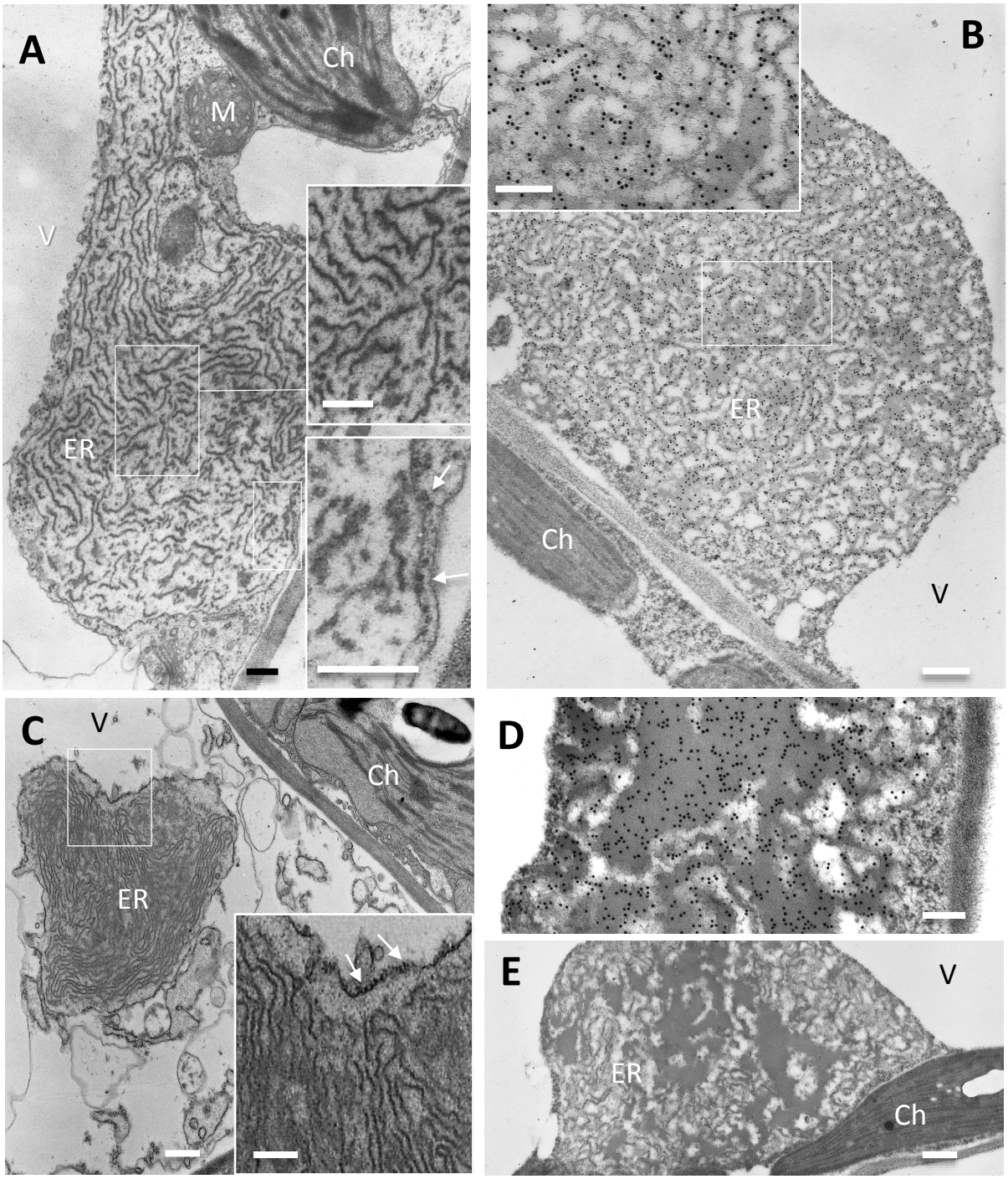


Fig. 7

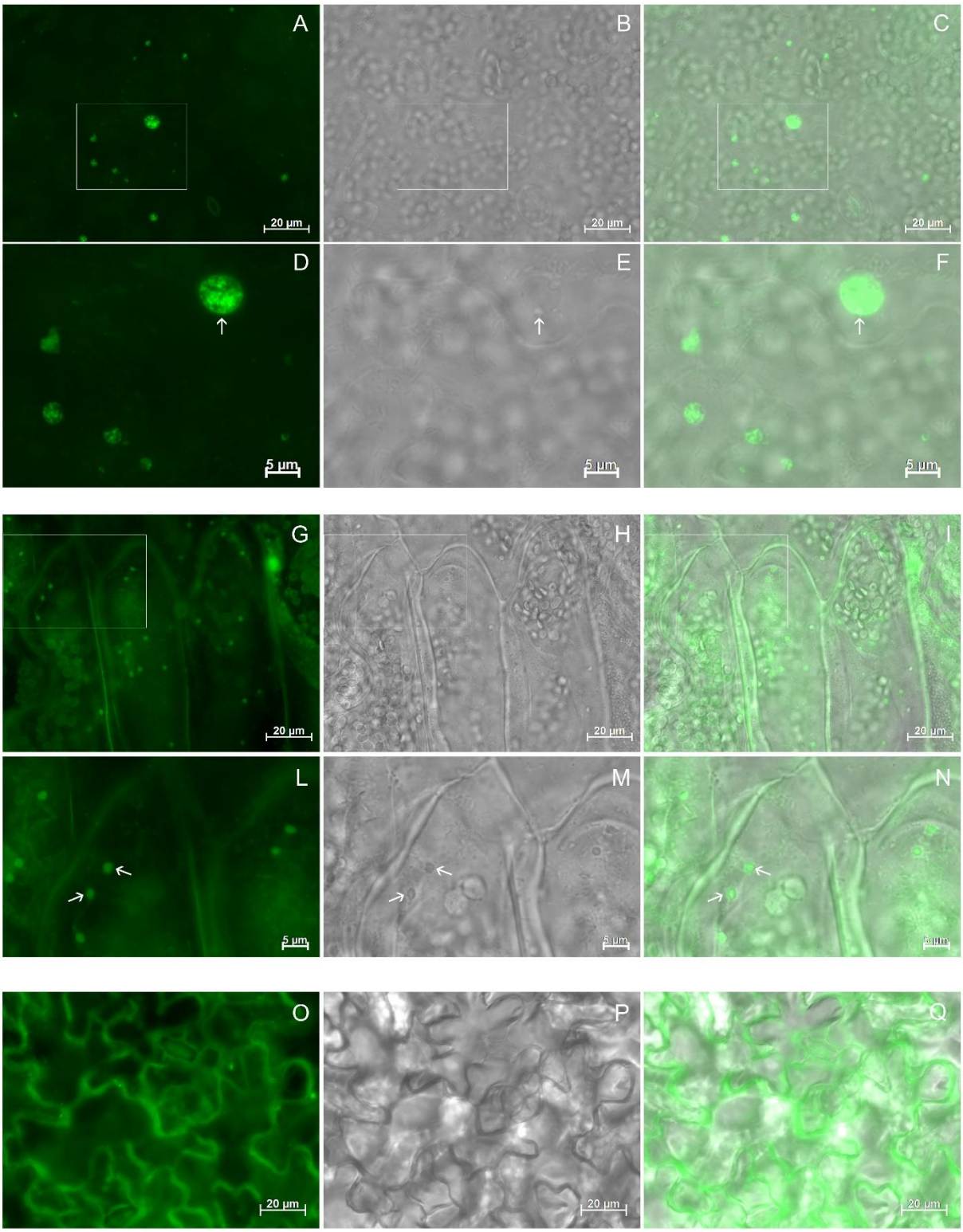


Fig. 8

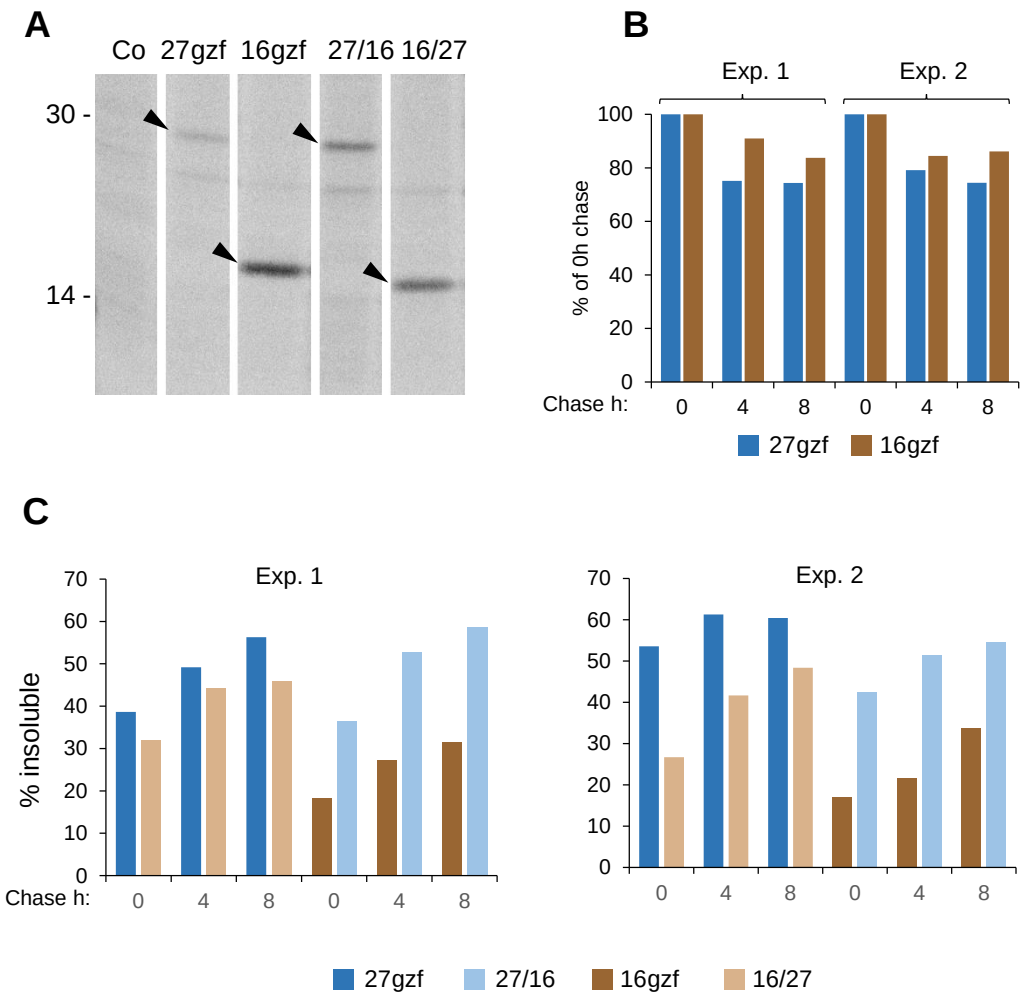


Fig. 9