A Family of NAI2-Interacting Proteins in the Biogenesis of the ER Body and Related Structures[1][OPEN]

Zhe Wang,[a,2] Xifeng Li,[a,b,2] Nana Liu,[a,c] Qi Peng,[a,d] Yuexia Wang,[a,e] Baofang Fan,[a] Cheng Zhu,[b,3] and Zhixiang Chen[a,b,3,4]

[a]Department of Botany and Plant Pathology and Center for Plant Biology, 915 W. State Street, Purdue University, West Lafayette, Indiana 47907-2054
[b]College of Life Science, China Jiliang University, Hangzhou 310018, China
[c]College of Science, China Agricultural University, Beijing 100193, China
[d]Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
[e]College of Life Sciences, Henan Agricultural University, Zhengzhou 450002, China

ORCID IDs: 0000-0001-9847-296X (Z.W.); 0000-0002-5933-3622 (X.L.); 0000-0001-9745-8525 (N.L.); 0000-0002-3593-2153 (Q.P.); 0000-0002-8125-3346 (B.F.); 0000-0002-8056-4943 (C.Z.); 0000-0002-5472-4560 (Z.C.).

Plants produce different types of endoplasmic reticulum (ER)-derived vesicles that accumulate and transport proteins, lipids, and metabolites. In the Brassicales, a distinct ER-derived structure called the ER body is found throughout the epidermis of cotyledons, hypocotyls, and roots. NAI2 is a key factor for ER body formation in Arabidopsis (*Arabidopsis thaliana*). Homologs of NAI2 are found only in the Brassicales and therefore may have evolved specifically to enable ER body formation. Here, we report that three related Arabidopsis NAI2-interacting proteins (NAIP1, NAIP2, and NAIP3) play a critical role in the biogenesis of ER bodies and related structures. Analysis using GFP fusions revealed that all three NAIPs are components of the ER bodies found in the cotyledons, hypocotyls, and roots. Genetic analysis with *naip* mutants indicates that they have a critical and redundant role in ER body formation. NAIP2 and NAIP3 are also components of other vesicular structures likely derived from the ER that are formed independent of NAI2 and are present not only in the cotyledons, hypocotyls, and roots, but also in the rosettes. Thus, while NAIP1 is a specialized ER body component, NAIP2 and NAIP3 are components of different types of ER-derived structures. Analysis of chimeric NAIP proteins revealed that their N-terminal domains play a major role in the functional specialization between NAIP1 and NAIP3. Unlike NAI2, NAIPs have homologs in all plants; therefore, NAIP-containing ER structures, from which the ER bodies in the Brassicales may have evolved, are likely to be present widely in plants.

The endoplasmic reticulum (ER) is an interconnected network of membrane sacs and tube-like cisternae found in eukaryotic cells. The ER is the gateway of intracellular trafficking of proteins to a variety of cellular destinations along the secretory pathway (Vitale and Denecke, 1999). Most membrane and soluble proteins that are synthesized and pass quality control in the ER move to the Golgi apparatus through the coat protein complex II-coated vesicles before transport to other endomembrane compartments or to the extracellular space (Benham, 2012). In all eukaryotes, this is the best characterized mechanism of the endomembrane system for transport of proteins synthesized on the ER. In plant cells, however, there are other specialized compartments derived from the ER with various sizes and shapes that contain proteins actively synthesized on the ER without traveling through the Golgi apparatus (Hara-Nishimura et al., 2004). Many of these ER-derived compartments travel to and are incorporated into vacuoles in a Golgi-independent manner. In plant seeds, for example, some of the ER-derived protein bodies, which contain a high amount of proteins synthesized on the ER, proceed directly to the protein storage vacuoles independent of the Golgi apparatus and other post-Golgi compartments in the secretory pathway (Chrispeels and Herman, 2000). Specialized ER-derived vesicles also play a role in the biogenesis of lytic vacuoles (Viotti et al., 2013). Therefore, plant cells are unique in the flexibility of the ER to assemble a variety of ER-derived compartments for direct transport to other destinations, particularly to the vacuoles.

[1]This work is supported by the U.S. National Science Foundation (grant no. IOS1456300 and IOS-1758767).
[2]These authors contributed equally to the article.
[3]Author for contact: zhixiang@purdue.edu.
[4]Senior author.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Zhixiang Chen (zhixiang@purdue.edu).
Z.C. conceived the original research plans; C.Z and Z.C. supervised the experiments; Z.W. and X.L. performed most of the experiments; N.L., Q.P., Y.W., and B.F. performed some of the experiments; Z.W., X.L., C.Z., and Z.C. analyzed the data and wrote the article with contributions of all the authors.
[OPEN]Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.18.01500

Among those specialized ER-derived compartments that have been extensively characterized are the ER bodies, which are produced only by plants in the Brassicales order, including Arabidopsis (*Arabidopsis thaliana*; Nakano et al., 2014). Unlike other ER-derived vesicles, the ER bodies are rod-shaped, approximately 1 μm in diameter and 10 μm in length, and can be observed in transgenic Arabidopsis plants expressing ER-targeted GFP (Hawes et al., 2001; Hayashi et al., 2001). Analysis using electron microscopy showed that the ER bodies contain a single membrane covered by ribosomes and are connected with ER tubules and cisternae, indicating that the ER bodies are continuous to the whole ER network (Hayashi et al., 2001). The major protein component of the ER bodies in Arabidopsis is PYK10/BGLU23, a β-glucosidase with a Lys-Asp-Glu-Leu ER retention signal at its C terminus (Matsushima et al., 2003). Two integral membrane proteins with a metal ion transporter activity, MEMBRANE OF ER BODY1 (MEB1) and MEB2, have been identified to accumulate specifically at the membranes of the ER bodies in Arabidopsis (Yamada et al., 2013). Genetic analysis has identified two genes, *NAI1* and *NAI2*, with an important role in the ER body formation in Arabidopsis (Matsushima et al., 2004; Yamada et al., 2008). *NAI1* encodes a basic helix-loop-helix-type transcription factor and functions as a master regulator of the ER body formation by regulating the expression of genes encoding PYK10, NAI2, MEB1, MEB2, and other related proteins (Matsushima et al., 2004). *NAI2* encodes an ER body component that determines the ER body formation in Arabidopsis (Yamada et al., 2008). In the *nai2* mutants, PYK10, MEB1, and MRB2 are diffused throughout the ER and the levels of PYK10 are reduced, indicating that NAI2 promotes accumulation of PYK10 by mediating the formation of the ER bodies (Yamada et al., 2008). NAI2 forms complexes with MEB1 and MEB2 and therefore may be responsible for the recruitment and organization of these ER body membrane proteins (Yamada et al., 2013). Homologs of NAI2 are found only in plants in the Brassicaceae order that form ER bodies, suggesting that NAI2 has a specific role in the formation of the ER-derived vesicles (Yamada et al., 2008).

The ER bodies are enriched in the cotyledons and hypocotyls of Arabidopsis seedlings and in roots of both seedlings and mature plants (Nakano et al., 2014). However, the number of the ER bodies in the rosette leaves of mature plants is very low but can be induced by wounding in a jasmonic-acid-dependent manner (Matsushima et al., 2002; Ogasawara et al., 2009). Recently, it has been reported that Arabidopsis TONSOKU (TSK)-ASSOCIATED PROTEIN1 (TSA1), a close homolog of NAI2, plays a critical role in jasmonic-acid-induced ER body formation (Geem et al., 2019). These observations suggest a possible role of the ER bodies in plant responses to pathogens, herbivores, and other stresses. This is supported by the recent finding that the abundant PYK10 β-glucosidase in the ER bodies has a myrosinase activity that hydrolyzes indole glucosinolates, thereby generating chemically reactive products toxic to pathogens and herbivores (Nakano et al., 2017). In addition, genes associated with the ER body, glucosinolate biosynthesis, and metabolism display a striking coexpression pattern, suggesting strong coordination among these processes (Nakano et al., 2017). Methylerythritol cyclodiphosphate, a precursor of plastidial isoprenoids and a stress-specific retrograde signaling metabolite, plays a key role in coordinately promoting the ER body formation and induction of indole glucosinolate metabolism through transcriptional regulation of the key regulators NAI1 and MYB51/122 transcription factors, respectively (Wang et al., 2017). The role of ER body formation has also been demonstrated in response of Arabidopsis plants to the beneficial fungus *Piriformospora indica*. In the *pyk10* and *nai1* mutants, infection by the beneficial fungus led to fungal overgrowth without beneficial effects on the plants (Sherameti et al., 2008). This suggests that ER body formation plays a role in plant defense that enables controlled fungal colonization to establish a mutualistic interaction between the symbiotic partners (Sherameti et al., 2008). The ER body may also play a role in plant responses to other stresses, including drought and metal ion toxicity (Yamada et al., 2013; Kumar et al., 2015).

While a substantial amount of information on the components, function, and regulation of the ER bodies has been generated, there are important unresolved questions on the biogenesis and evolution of the ER bodies. ER body formation is likely to be a complex process involving proteins in addition to NAI2. It is unclear whether the ER bodies have originated in the Brassicales or evolved from related ER-derived vesicles. In this study, we report identification and characterization of three closely related NAI2-interacting proteins (NAIP1, NAIP2, and NAIP3) from Arabidopsis. Genetic analysis revealed that formation of NAI2- or PKY10-containing ER bodies was normal in the *naip1*, *naip2*, and *naip3* single and double mutants but greatly reduced in the triple mutant. Consistent with their critical and redundant role in the ER body formation, all three NAIP proteins are components of ER bodies based on analysis using GFP fusion proteins. However, while NAIP1 is only associated with the ER bodies, NAIP2 and NAIP3 are also associated with other types of ER-derived vesicles that are formed in an NAI2-independent manner not only in the cotyledons, hypocotyls, and roots, but also in the rosette leaves of Arabidopsis plants. Functional dissection of the NAIP proteins through analysis of chimeric proteins provided further insights into the structural domains with a role in determining the characteristics of these ER-derived compartments. Unlike NAI2, homologs of NAIP proteins are found in all plants; therefore, NAIP-containing ER structures are likely to be present widely in plants. Based on these results, we propose that the ER bodies from the Brassicales with a specialized function may have evolved from some of the NAIP-containing ER

compartments that are likely to be present widely in plants.

RESULTS

Identification and Structural Characterization of Proteins That Interact with NAI2

We have recently reported identification of three related AUTOPHAGY-RELATED PROTEIN8 (ATG8)-interacting proteins, ATI3a, ATI3b, and ATI3c, from Arabidopsis and provided evidence that these ATI3 proteins function as selective autophagy receptors through interacting with two ER-localized UBIQUITIN ASSOCIATED DOMAIN CONTAINING2 (UBAC2) proteins and target degradation of specific unknown ER components during plant stress responses (Zhou et al., 2018). To further study the ATI3/UBAC2-mediated selective autophagy pathway, we attempted to identify UBAC2-interacting proteins using yeast two-hybrid screens. Using UBAC2a as a bait, we screened an Arabidopsis cDNA library and identified a number of UBAC2-interacting clones, among which were two that correspond to At4g15545. Quantitative assays of β-galactosidase activity for *LacZ* reporter gene expression revealed that the interaction of UBAC2 and the protein encoded by At4g15545 was relatively weak and the interaction could not be confirmed reproducibly by bimolecular fluorescence complementation (BiFC) assays in *Nicotiana benthamiana*. Using the protein encoded by At4g15545 as a bait, we identified strong positive clones from yeast two-hybrid screens that encode the C-terminal 125-amino acid residues of Arabidopsis NAI2, an ER body component found only in the Brassicaceae (Yamada et al., 2008). This NAI2-interacting protein encoded by At4g15545 was named NAIP1.

NAIP1 is a protein of 337 amino acids with two homologs in Arabidopsis: NAIP2 (At1g16520) and NAIP3 (At1g56080), with 325 and 310 amino acids, respectively (Supplemental Fig. S1A). The NAIP proteins are most conserved at their C-terminal domains (Supplemental Fig. S1A), which are highly homologous to the N-terminal protein-binding module of harmonin, also known as harmonin homology domain (HHD; Supplemental Fig. S1B). The three proteins also shared a high degree of sequence similarity at the N-terminal domains, which are predicted to display coiled-coil (CC) structures (Supplemental Fig. S1). At the middle section, the three proteins are highly divergent in amino acid sequences but share five Thr or Ser residues immediately preceding a Pro residue (TP or SP; Supplemental Fig. S1), which are the major regulatory phosphorylation motifs by a large family of so-called Prodirected protein kinases, including cyclin-dependent protein kinases and mitogen-activated protein kinases (Lee et al., 2005). Thus, the NAIP proteins are rich in protein-interacting motifs and are potentially regulated by protein phosphorylation.

To determine whether NAIP2 and NAIP3 also interact with NAI2, we cloned them into the yeast pBD-Gal4 vector and tested the interactions using yeast two-hybrid assays. As shown in Figure 1, both NAIP2 and NAIP3 interacted with NAI2, and quantitative assays of β-galactosidase activity for the *LacZ* reporter gene expression revealed that their interaction with NAI2 was even stronger than NAIP1's interaction with NAI2 (Fig. 1). NAIP1 also interacted with itself and with NAIP2 and NAIP3 (Fig. 1). Using truncated N-terminal and C-terminal domains of NAIP1 as baits, we discovered that the N-terminal CC domain of NAIP1 was sufficient for interaction with itself (Fig. 1). By contrast, the C-terminal HHD, but not the N-terminal CC domain of NAIP1, interacted with NAI2 in yeast cells (Fig. 1). Therefore, the N-terminal CC domains of the three NAIP proteins mediated their self and mutual

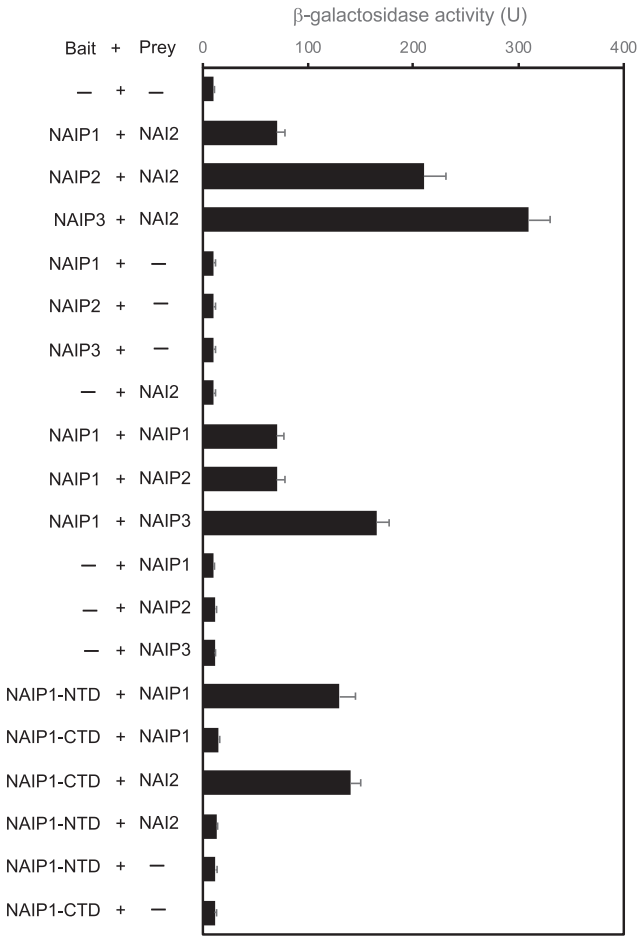

Figure 1. Interaction of NAIP proteins with NAI2 and self-interactions in yeast cells. The binding domain fusion (bait) of a gene for a whole NAIP protein or the N-terminal CC domain (NAIP-NTD) or the C-terminal HHD domain (NAIP-CTD) of a NAIP protein was cotransformed with the activation domain fusion (prey) of *NAI2* or a *NAIP* gene into yeast cells. Proteins were isolated from the yeast cells and assayed for β-galactosidase activity using *o*-nitrophenyl-β-D-galactopyranose (ONPG) as substrate. Data represent means and SES ($n = 5$).

interactions, while their C-terminal HHD domain is responsible for their interaction with NAI2.

To determine whether NAIP and NAI2 proteins interact in plant cells, we performed BiFC in Arabidopsis. We fused the three Arabidopsis NAIP proteins to the N-terminal yellow fluorescent protein (YFP) fragment (NAIP-N-YFP) and NAI2 to the C-terminal YFP fragment (NAI2-C-YFP). As a control, we also fused PYK10, an Arabidopsis β-glucosidase located in the ER bodies and often used as an ER body marker (Matsushima et al., 2003, 2004; Yamada et al., 2008), to the C-terminal YFP fragment (PYK10-C-YFP). These constructs were transformed into Arabidopsis, and transgenic plants were crossed to generate lines coexpressing appropriate BiFC constructs for fluorescence complementation. When fused NAIP-N-YFP was coexpressed with NAI2-C-YFP in Arabidopsis cotyledons, BiFC signals were detected in transformed cells (Fig. 2). While the BiFC fluorescent signals resulting from the interactions between the NAIP and NAI2 proteins were largely dispersed, we observed a substantial number of punctate structures of 1–3 μm in diameter with all three pairs of the vectors (Fig. 2). Control BiFC experiments in which NAIP1-N-YFP was coexpressed with fused PYK10-C-YFP or unfused C-YFP, or unfused N-YFP was coexpressed with NAI2-C-YFP, did not show fluorescence (Fig. 2). Failure to observe fluorescent signals in the control BiFC experiments was not due to lack of expression of the fused proteins, as indicated by western blotting (Supplemental Fig. S2). We also fused PYK10 to the N-terminal YFP fragment (PYK10-N-YFP), coexpressed with NAI2-C-YFP, but failed to observed BiFC fluorescent signals (Supplemental Fig. S3). To determine the cellular nature of the BiFC fluorescent signals, we also crossed the transgenic plants coexpressing NAIP-N-YFP and NAI2-C-YFP with a transgenic line expressing a PYK10-mCherry construct. The yellow fluorescent signals from the complementation between the NAIP-N-YFP and NAI2-C-YFP proteins were generally colocalized with those of the PYK1-mCherry signals (Supplemental Fig. S4).

The three NAIP proteins contain no predicted transmembrane domain or signal peptide (Supplemental Fig. S1A) but interact with NAI2 at the ER bodies based on their colocalization with the PKY10 ER body marker in the cotyledons of Arabidopsis plants (Supplemental Fig. S4). Therefore, the three Arabidopsis NAIP proteins are likely to be localized on the cytosolic side of the ER, ER bodies, or other membrane structures. On the other hand, both NAI2 and its close homolog TSA1 contain a signal peptide at their N terminus and have been localized to the ER bodies, suggesting that they are both luminal components of the ER bodies. This raises a topology problem of how a luminal protein such as NAI2 interacts directly with the NAIP proteins on the cytosolic face of the ER bodies in plant cells. Interestingly, several studies have reported that both NAI2 and TSA1 interact with proteins that are not luminal proteins of the ER or ER bodies. TSA1 interacts with TSK, a protein with an important role in cell division (Suzuki et al., 2005). TSA1 also interacts with CSN1, one of the

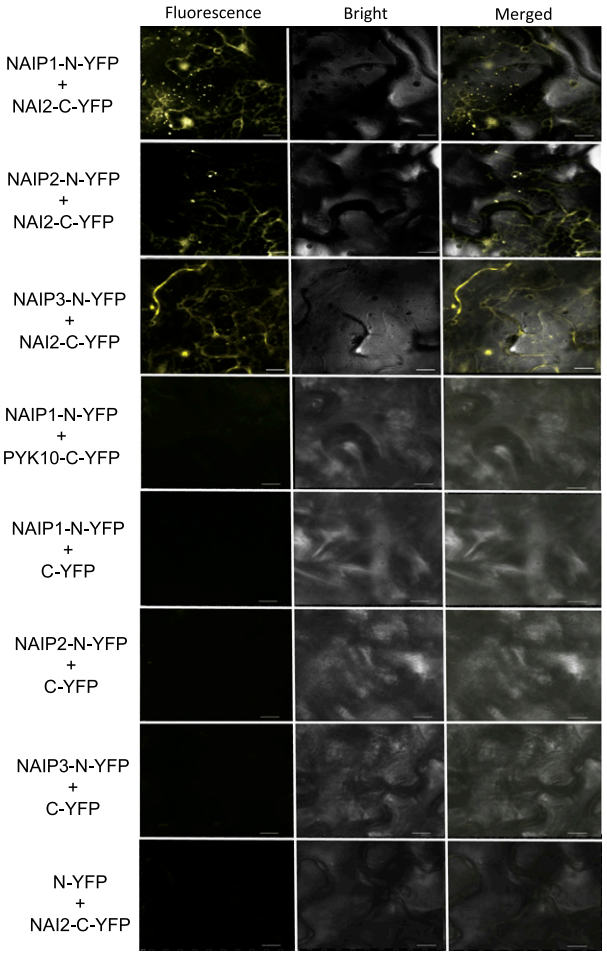

Figure 2. BiFC assays of NAIP-NAI2 interactions in Arabidopsis cotyledons. BiFC fluorescence was observed in the epidermal cells of the cotyledons in Arabidopsis, from complementation of the N-terminal half of the YFP fused with NAIP1, NAIP2, and NAIP3 (NAIP1-, NAIP2-, and NAIP3-N-YFP, respectively) with the C-terminal half of the YFP fused with NAI2 (NAI2-C-YFP). No fluorescence was observed when NAIP-N-YFP was coexpressed with fused PYK10-C-YFP or unfused C-YFP or when unfused N-YFP was coexpressed with NAI2-C-YFP. YFP epifluorescence, bright-field, and merged images of the same cells are shown. Bar, 10 μm.

subunits of the COP9 signalosome, which participates in diverse cellular and developmental processes (Li et al., 2011). NAI2 interacts with AFL1, a peripheral membrane protein associated with both plasma membrane and endomembranes (Kumar et al., 2015). The interactions of NAI2 and TSA1 with these proteins were demonstrated not only by yeast two-hybrid assays but also by BiFC and coimmunoprecipitation (Figs. 2 and 3; Supplemental Fig. S4; Suzuki et al., 2005; Li et al., 2011; Kumar et al., 2015). The functional relevance of their interactions has also been established through genetic analysis (Suzuki et al., 2005; Li et al., 2011; Kumar et al., 2015). The demonstrated interactions of both NAI2 and TSA1 with a variety of proteins that are not luminal proteins of the ER and ER bodies suggest that NAI2 and

Figure 3. Promoter-GUS activities. Arabidopsis plants were transformed with the GUS reporter gene driven by the *NAIP1*, *NAIP2*, or *NAIP3* promoter. GUS staining was performed in seedlings (top), rosette leaves of mature plants (middle), and inflorescence of flowering plants (bottom). At least five independent transgenic lines for each construct were used in the assays with very similar results.

TSA1 may not be fully luminal. Indeed, it has been predicted that TSA1 contains a transmembrane domain between the N-terminal Glu-Phe-Glu repeats and the C-terminal TSK-interacting domain (Suzuki et al., 2005). This information and the demonstrated interactions with proteins on the cytosolic side of the ER bodies make it likely that both NAI2 and TSA1 are transmembrane proteins with a topology of their N-terminal Glu-Phe-Glu repeats in the luminal of the ER bodies and their C-terminal domain on the cytosolic side of the ER bodies. This topology is consistent with the recent demonstration of N-glycosylation at the N terminus of both NAI2 and TSA1 (Geem et al., 2019). It is also consistent with the fact that both TSK and CSN1 interact with the C-terminal domain of TSA1, which is predicted to be cytosolic (Suzuki et al., 2005; Li et al., 2011). Likewise, the NAIP proteins interact with the C-terminal 125-amino acid residues of NAI2. To provide additional evidence for the topology of NAI2, we generated transgenic Arabidopsis plants expressing NAI2 and PYK10 that were tagged with a myc epitope at their C terminus. PYK10 is a fully luminal protein of the ER bodies. Microsomal fractions were isolated from the seedlings of the transgenic plants and were analyzed by protease K protection assays. By incubation of the microsomal fractions containing the NAI2- or PYK10-myc proteins with the protease, we found that NAI2-myc was more rapidly degraded than PYK10-myc by protease K (Supplemental Fig. S5). However, increasing the time of the protease treatment also led to degradation of PYK10 (Supplemental Fig. S5), which could be due to the breakdown of the ER bodies, which may be prone to rupture (Nakano et al., 2014). Incubation of the microsomal fractions with a detergent during the protease digestion abolished the difference in degradation kinetics between NAI2 and PYK10 (Supplemental Fig. S5).

Expression of *NAIP* Genes

To determine the expression patterns of the genes encoding the three NAIP proteins, we generated transgenic Arabidopsis plants harboring the *GUS* reporter gene under control of the *NAIP* gene promoters. The promoter activities of the *NAIP* genes were determined by GUS activity staining in the seedlings, rosette leaves, and inflorescence. In seedlings, *NAIP1* was expressed at high levels in the cotyledon, hypocotyl, and small rosette leaves but at low levels in the roots (Fig. 3). *NAIP1* was also expressed at high levels in the rosette leaves of mature plants and in the inflorescence of flowering plants (Fig. 3). On the other hand, *NAIP2* expression in seedlings appeared to be largely restricted to the vascular tissues of shoots, but its expression was relatively high in roots (Fig. 3). In the rosette leaves of mature plants, *NAIP2* expression was low and again mostly limited to leaf veins (Fig. 3). Expression of *NAIP2* was also detected in the inflorescence of flowering plants, mostly limited to the main branch (Fig. 3). The expression patterns of *NAIP3* were similar to those of NAIP1, albeit at reduced levels. In seedlings, *NAIP3* expression was also detected in the cotyledons, hypocotyls, and small rosette leaves but at very low levels in the roots. The expression levels of *NAIP3* in the cotyledon of seedlings and in the rosette leaves of mature plants were substantially higher in the vascular veins than the surrounding tissues (Fig. 3). The higher expression levels of *NAIP3* in the vascular tissues were also apparent in the inflorescence, including flowers, of flowering plants (Fig. 3).

Subcellular Localization of NAIP Proteins in Arabidopsis

To investigate the subcellular localization of the NAIP proteins in plant cells, we coexpressed the NAIP-GFP constructs with a PYK10-mCherry fusion construct

in Arabidopsis (Fig. 4A). In the cotyledons of these transgenic plants, we observed that the punctuate fluorescent structures of PYK10-mCherry were mostly colocalized with those from coexpressed NAIP1- and NAIP2-GFP (Fig. 4). On the other hand, only about 50% of the punctate signals of NAIP3-GFP were colocalized with those of the coexpressed PYK10-mCherry (Fig. 4). These results suggest that the three NAIP proteins are all components of the ER bodies, but some of them, such as NAIP3, may also be components of other types of ER-derived vesicular structures.

Role of NAI2 in the Formation of NAIP-Containing Vesicular Structures

To further determine the relationship of the NAIP-containing vesicular structures, the ER bodies, and other ER-derived structures, we introduced the NAIP-GFP fusion constructs into both wild-type and *nai2* mutant plants. In wild type, we observed that expression of NAIP1-GFP resulted in formation of punctate GFP signals in the cells of cotyledons, hypocotyls, and roots, but not in the cells of rosette leaves (Fig. 5). This pattern of tissue-specific formation of punctate NAIP1-GFP structures is identical to those of the NAI2 and PKY10 ER body markers (Matsushima et al., 2003; Yamada et al., 2008). Furthermore, in the *nai2* mutant, the numbers of punctate NAIP1-GFP fluorescent structures in the cotyledons, hypocotyls, and roots were greatly reduced (Fig. 5; Supplemental Fig. S6). Both the tissue specificity and the NAI2 dependency strongly suggest that the punctate NAIP1-GFP structures observed in transgenic Arabidopsis plants are the ER bodies.

There were relatively fewer punctate GFP signals in wild-type transgenic plants expressing NAIP2-GFP than those expressing NAIP1-GFP (Fig. 5; Supplemental Fig. S6). However, the few punctate GFP signals were observed in the cells of not only the cotyledons, hypocotyls, and roots, but also the rosette leaves of both wild-type and *nai2* mutant plants (Fig. 5; Supplemental Fig. S6). On the other hand, expression of NAIP3-GFP generated a large number of punctate fluorescent structures in the

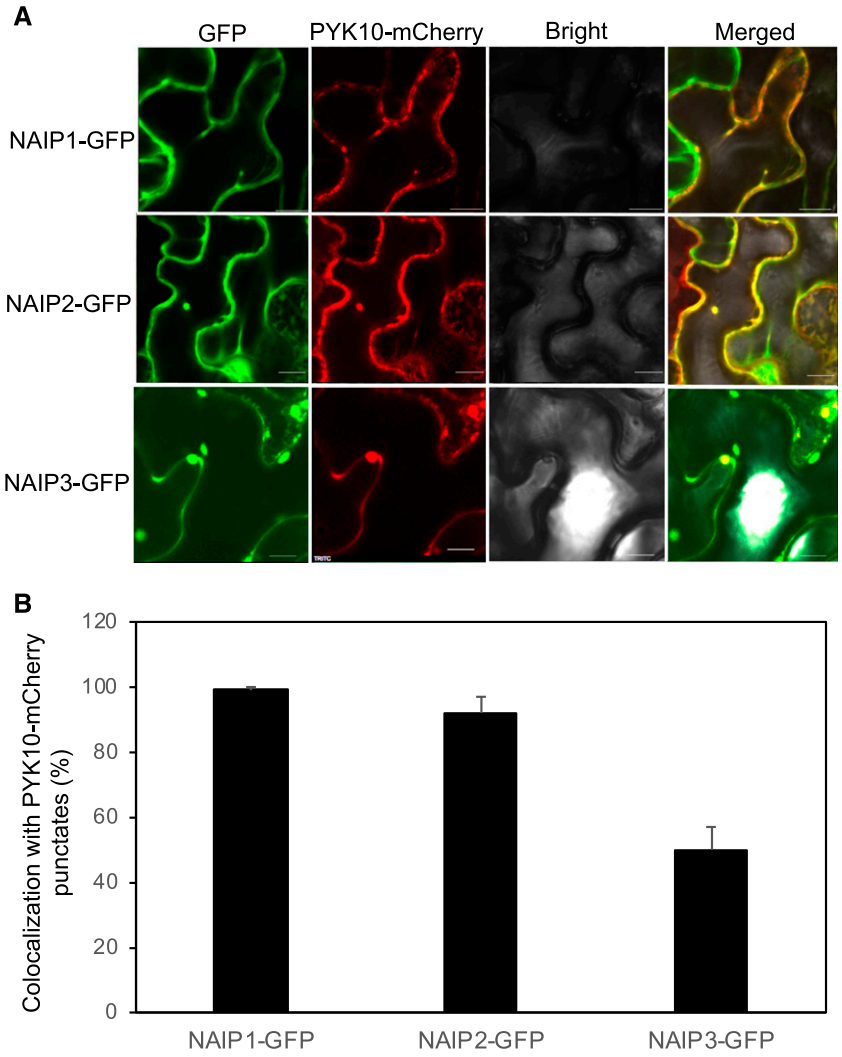

Figure 4. Colocalization of NAIP proteins with PYK10 in Arabidopsis cotyledon. A, Representative images of confocal fluorescence microscopy analysis in the cotyledons. Bar, 10 μm. Transgenic Arabidopsis plants expressing the genes for the NAIP-GFP fusion proteins were crossed with those expressing the PYK10-mCherry ER body marker and examined for colocalization in the F1 progeny. B, Percentages of colocalized NAIP-GFP and PYK10-mCherry punctate fluorescent signals. The means and standard deviations were calculated from 50 sections (100 $\times$ 100 μm) from five plants.

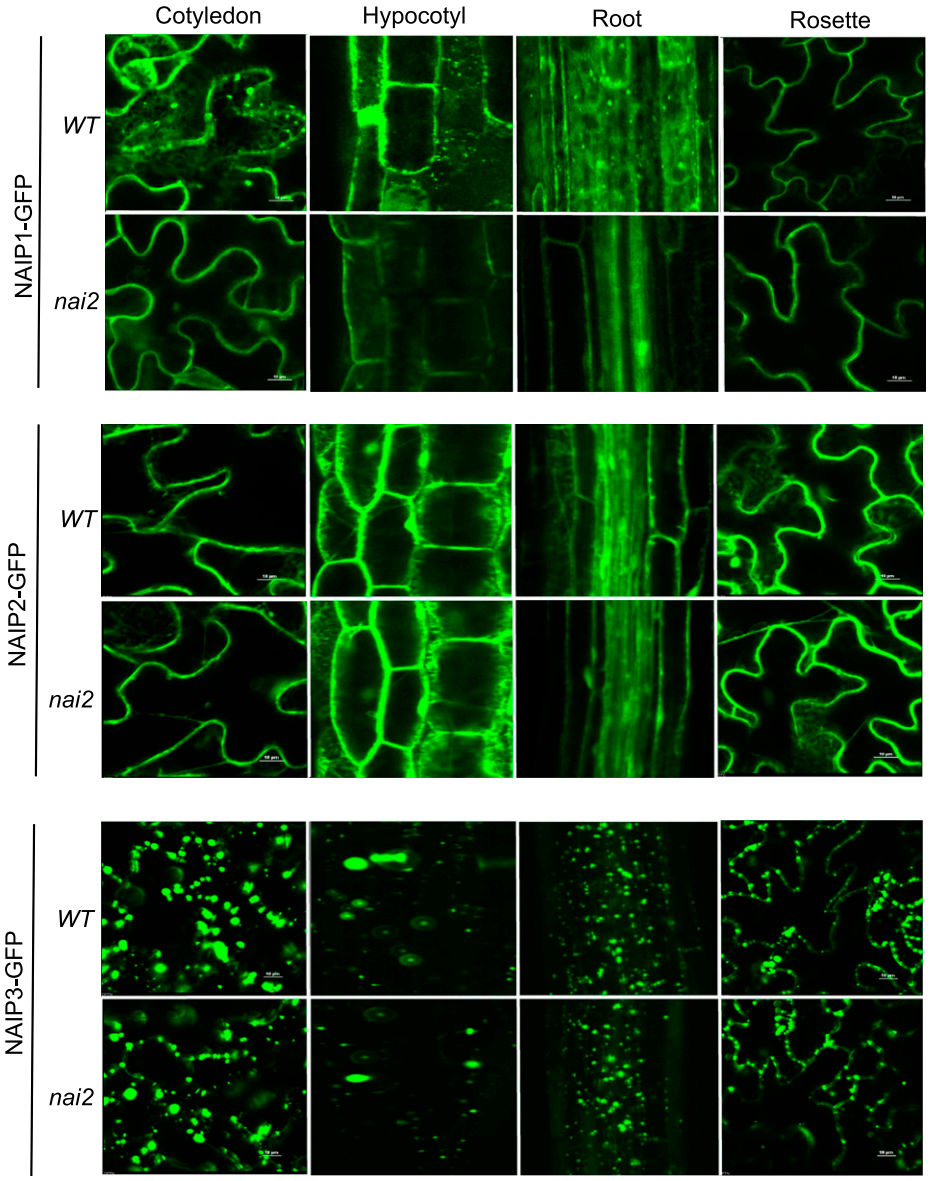

Figure 5. Formation of NAIP1-, NAIP2-, and NAIP3-containing vesicles in the *nai2* mutant. The constructs for the NAIP1-, NAIP2-, and NAIP3-GFP fusion proteins were transformed into both wild-type (WT) and *nai2* mutant plants. Confocal fluorescence microscopy analysis was performed in the cotyledons, hypocotyls, and roots of seedlings and the rosette leaves of mature plants. At least 10 independent transgenic lines for each construct and each genotype were used in the assays with very similar results. Bar, 10 μm.

cells of the cotyledons, hypocotyls, roots, and rosette leaves in both wild-type and *nai2* mutant plants (Fig. 5; Supplemental Fig. S6). Thus, unlike NAIP1, NAIP2, and NAIP3 are associated not only with the ER bodies but also with other vesicular structures whose formation is ubiquitous and NAI2 independent.

Size Difference among NAIP-Containing Vesicular Structures

Besides the tissue specificity and NAI2 dependency, we observed differences in the range of sizes of these fluorescent vesicular structures containing different NAIP proteins. In cotyledons, NAIP1-containing vesicular structures have a mean diameter of 0.9512 μm and an SD of 0.2906 μm (Figs. 5 and 6). Thus, the diameters of NAIP1-containing structures are very similar to the ER body's diameters of about 1 μm (Matsushima et al., 2003). The mean and SD of the diameters of the NAIP2-labeled structures in the cotyledons of wild-type plants were very similar to those of NAIP1-labeled structures (Figs. 5 and 6). On the other hand, the mean of the diameters of NAIP3-containing structures was 1.5259 μm, which were substantially larger than those of NAIP1- and NAIP2-labeled structures (Figs. 5 and 6). More notably, the variation of the sizes of the NAIP3-containing structures in wild-type cotyledons as indicated by an SD of 0.9711 μm in their diameters was substantially larger than those of NAIP1- and NAIP2-containing structures (Figs. 5 and 6). Therefore, NAIP3-containing vesicular bodies are highly heterogeneous, while NAIP1- and NAIP2-containing structures are more similar in size. A similar difference in size between NAIP1- and NAIP3-containing structures was also observed in hypocotyls and roots (Fig. 5).

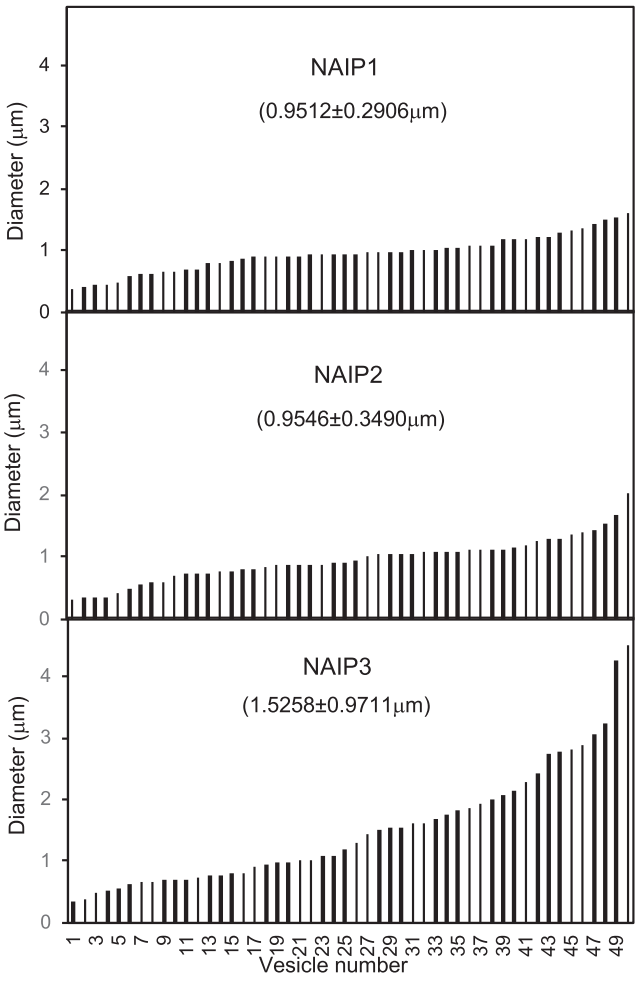

Figure 6. Size distribution of punctate NAIP-GFP signals in the cotyledon cells of transgenic plants. Confocal fluorescence microscopy analysis was performed in the cotyledons of transgenic plants expressing NAIP1-, NAIP2-, and NAIP3-GFP fusion proteins. The diameters of 50 punctual fluorescence signals in representative areas of cotyledon cells were determined, and the means and standard deviations ($n = 50$) are shown for each protein. For the rod-shaped punctual signals, the widths were used as their diameters.

Role of NAIP Proteins in the ER Body Formation

To determine the role of the three NAIP proteins in the formation of the ER bodies, we isolated transfer DNA (T-DNA) mutants for the three genes (Supplemental Fig. S7A). Quantitative reverse transcription-PCR indicated that these mutants accumulated few transcripts for the respective genes and are likely to be knockout mutants (Supplemental Fig. S7B). Through genetic crossing, we also generated *naip1/naip2* double and *naip1/naip2/naip3* triple mutants for the *NAIP* genes. The *naip* single, double, and triple mutants are indistinguishable from wild type in morphology, growth, and development. To examine the formation of the ER bodies in the mutants, we introduced both the NAI2-mCherry and PKY10-mCherry constructs into Arabidopsis wild-type and *naip* mutants. As expected, we observed NAI2- and PYK10-labeled ER bodies in the cells of cotyledons, hypocotyls, and roots, but not in the cells of rosette leaves in wild-type plants (Fig. 7). Similar numbers of NAI2- and PYK10-labeled ER bodies were also observed in the *naip1*, *naip2*, and *naip3* single mutants and in the *naip1/niap2* double mutants (Supplemental Fig. S8). In the *naip1/naip2/naip3* triple mutant, however, the numbers of NAI2- and PYK10-labeled ER bodies were greatly reduced (Fig. 7). The extent of reduction of the PYK10-labeled ER bodies in the *naip1/naip2/naip3* triple mutant was similar to that in the *nai2* mutant (Fig. 7B). To confirm that the drastic reduction of the PYK10-labeled ER bodies in the *naip1/naip2/naip3* triple mutant was due to the mutations of the *NAIP* genes, we introduced the *NAIP1-*, *NAIP2-*, and *NAIP3-GFP* constructs into the triple mutant and found that they could fully restore the formation of the PYK10-labeled ER bodies (Supplemental Fig. S9). These results indicated that the three *NAIP* genes play an important and redundant role in the formation of the ER bodies in Arabidopsis. By contrast, formation of NAI2- and PYK10-labeled ER bodies was normal in the *ubac2a/ubac2b* double mutant (Supplemental Fig. S8).

Dissection of NAIP Protein Domains in Formation of ER-Derived Structures

Even though the three Arabidopsis NAI2-interacting proteins have similar domain architectures and play a redundant role in the formation of ER bodies, they differ in the ability to associate with distinct types of ER-derived structures that differ in tissue specificity, NAI2 dependency, and size distribution (Figs. 5 and 6). These differences were observed even when these genes were all expressed constitutively under the *Cauliflower mosaic virus* (*CaMV*) *35S* promoter and therefore result from their distinct protein structures. To test this possibility, we generated six chimeric NAIP proteins by swapping the three domains between NAIP1 and NAIP3 (Fig. 8A). The GFP fusion constructs for the chimeric NAIP genes under the *CaMV 35S* promoter were introduced into Arabidopsis and their punctate GFP signals in the cells of cotyledons and rosette leaves were compared to those of transgenic NAIP1-GFP and NAIP3-GFP plants. As shown in Figure 8B, replacing either the C-terminal HHD domain or the middle domain of NAIP1 with the corresponding domains of NAIP3 (chimeric 1-1-3 and 1-3-1) still led to formation of punctate GFP signals in the cotyledons but not in the rosette leaves. By contrast, replacing the N-terminal domain of NAIP1 with the corresponding domain of NAIP3 (chimeric 3-1-1) led to formation of punctate GFP signals in the cells of both cotyledon and rosette (Fig. 8B). On the other hand, replacing any single domain of NAIP3 with the corresponding domain from NAIP1 (chimeric 3-3-1, 3-1-3, and 1-3-3) did not alter the formation of punctate GFP structures in both cotyledon and rosette (Fig. 8B). Thus, either replacing the

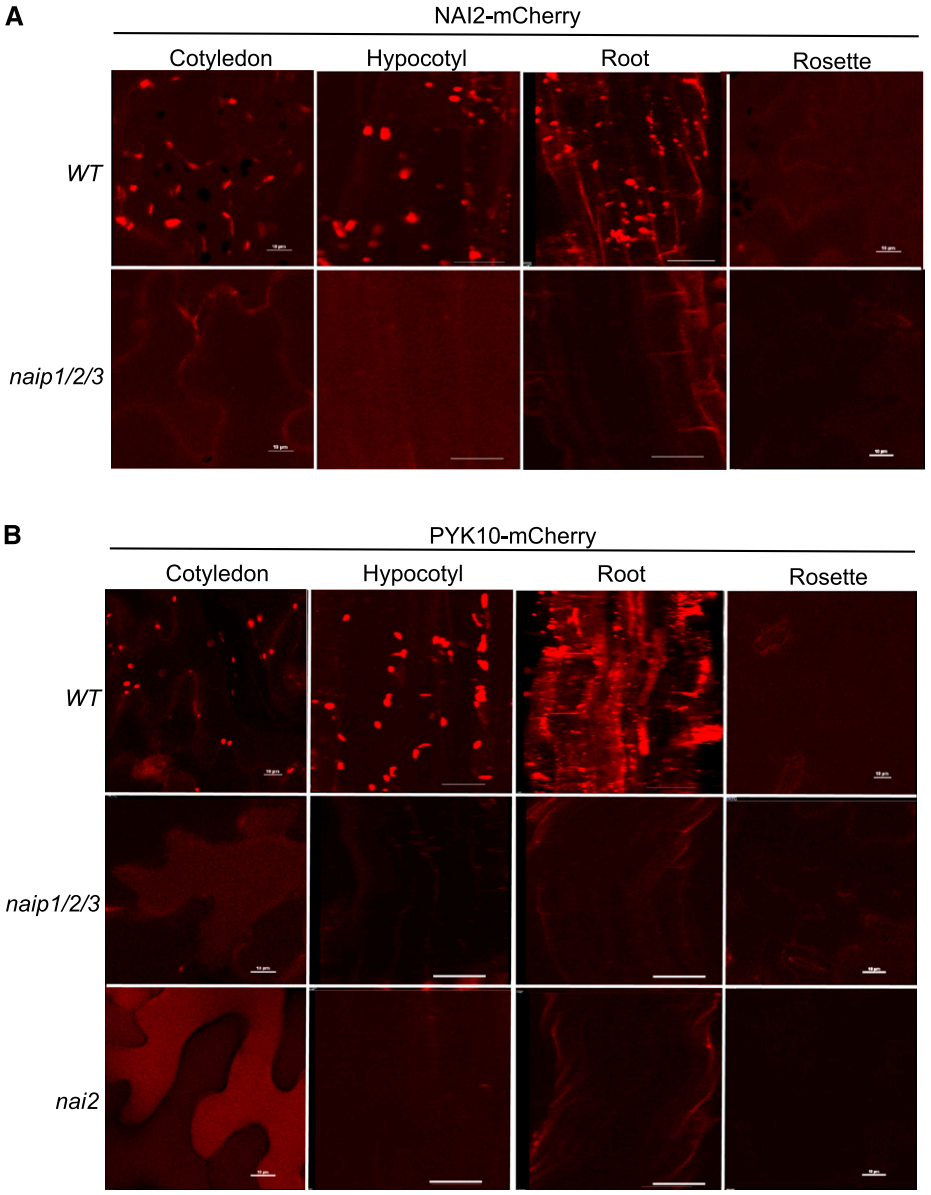

Figure 7. ER body formation in the *naip1/naip2/naip3* triple mutant. A, Formation of the ER bodies in transformed wild-type (WT) and *naip1/naip2/naip3* triple-mutant plants using the NAI2-mCherry fusion protein as the ER body marker. B, Formation of the ER bodies in transformed wild-type, *naip1/naip2/naip3* triple-mutant, and *nai2* mutant plants using the PYK10-mCherry fusion protein as the ER body marker. The constructs for the NAI2-and PYK10-mCherry ER body markers were transformed into the indicated genotypes. Confocal fluorescence microscopy analysis was performed in the cotyledons, hypocotyls, and roots of seedlings and in the rosette leaves of mature plants. At least five transgenic lines for each construct and each genotype were used in the experiments with very similar results. Bar, 10 μm.

N-terminal CC domain or replacing both the middle and C-terminal HHD domains of NAIP1 with the corresponding domains of NAIP3 could lead to formation of NAIP-labeled structures not only in the cotyledons but also in the rosette levels.

We also compared the size difference among the ER-derived structures labeled by NAIP1, NAIP3, and their chimeric proteins. In cotyledons, we observed that the chimeric proteins containing the N-terminal CC domain from NAIP1 (chimeric 1-1-3, 1-3-1, and 1-3-3) all formed vesicular structures similar to those of NAIP1 (chimeric 1-1-1) with relatively small average diameters and small standard deviations (Fig. 8, B and C). On the other hand, the three chimeric NAIP proteins containing the N-terminal CC domain from NAIP3 (chimeric 3-3-1, 3-1-3, and 3-1-1) all formed vesicular structures similar to those of NAIP3 (chimeric 3-3-3) with relatively large values in both the average diameters and standard deviations (Fig. 8, B and C). In the rosette leaves, only NAIP3 and four chimeric NAIP proteins (chimeric 3-3-1, 3-1-3, 3-1-1, and 1-3-3 proteins) were able to form vesicular structures (Fig. 8B). Among the four chimeric proteins, three containing the N-terminal CC domain from NAIP3 (chimeric 3-3-1, 3-1-3, and 3-1-1) all formed puncture structures similar to those of NAIP3 (the 3-3-3 protein) with relatively large values in both the average diameters and standard deviations (Fig. 8, B and D). By contrast, the only construct containing the N-terminal CC domain from NAIP1 (chimeric 1-3-3) formed punctual signals with very small and relatively uniform sizes (Fig. 8, B and D). Thus, in both cotyledons and rosette leaves, the N-terminal domains of the NAIP proteins play a critical role in determining the size of the vesicular structures that are formed.

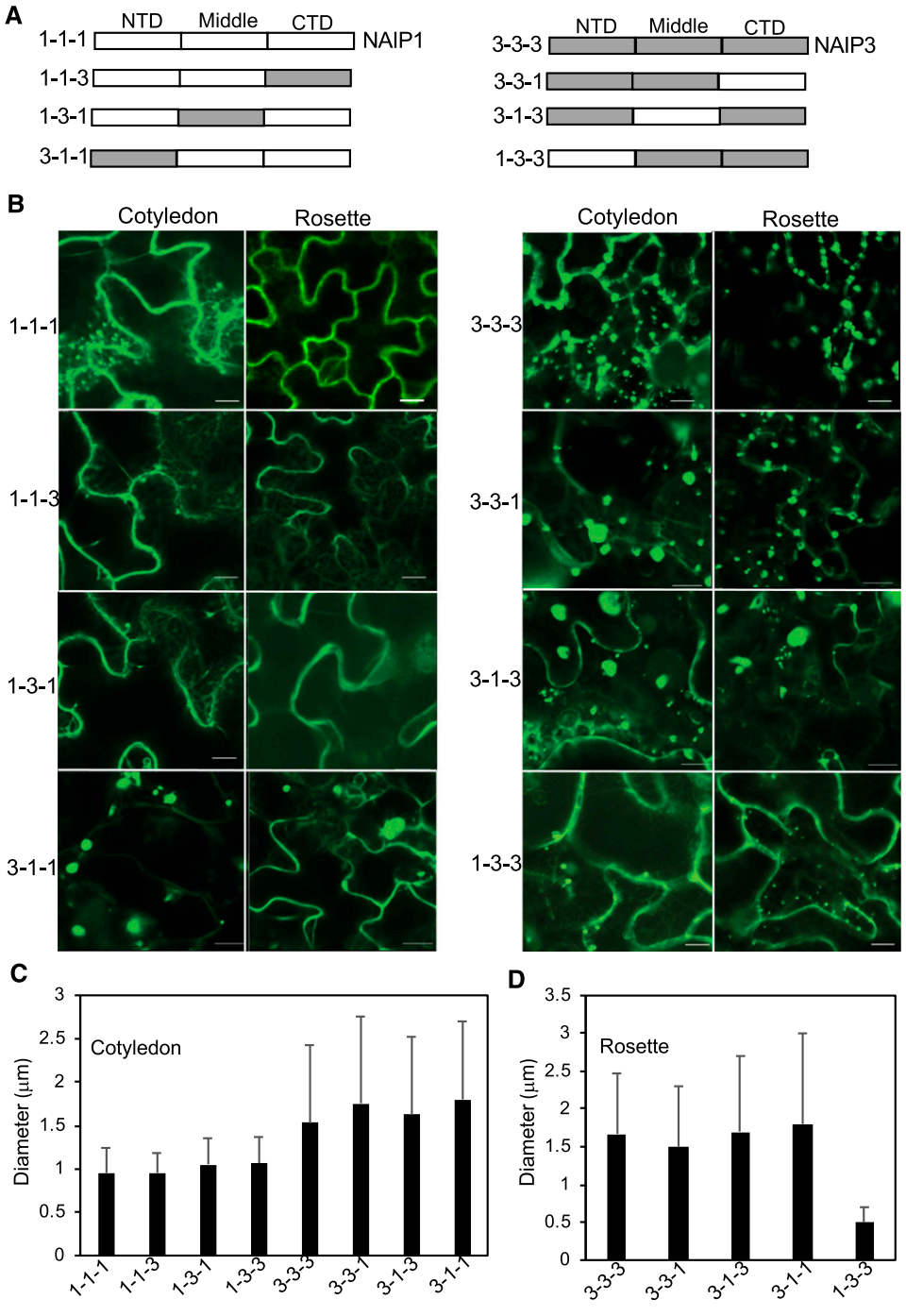

Figure 8. Functional analysis of the three structural domains of the NAIP proteins using domain swapping between NAIP1 and NAIP3. A, Schematic representation of NAIP1, NAIP3, and their six chimeric proteins. Like the original NAIP1 and NAIP3, each chimeric protein contains an N-terminal CC domain (NTD), a middle variable domain (middle), and a C-terminal HHD domain (CTD) and is given a three-digit name. The numbers 1 and 3 in a three-digit name refer to NAIP1 and NAIP3, respectively, as the sources of the domains in the chimeric protein. NAIP1 and NAIP3 are also denoted by the three-digit names 1-1-1 and 3-3-3, respectively. B, The GFP fusion genes for NAIP1, NAIP3, and six chimeric proteins were transformed into Arabidopsis wild-type plants. Confocal fluorescence microscopy analysis was performed in the cotyledon of seedlings and the rosette leaves of mature plants. At least five transgenic lines for each construct were used in the experiments with very similar results. Bar, 10 μm. C, Size distribution of punctual fluorescent signals in the cotyledon cells of transgenic plants expressing the GFP fusion genes for NAIP1, NAIP3, and six chimeric proteins. The means and standard deviations were calculated from at least 100 punctual signals in representative areas of cotyledon cells from five independent lines for each construct. D, Size distribution of punctual fluorescent signals in the rosette cells of transgenic plants expressing the GFP fusion genes for NAIP3 and four other chimeric proteins. The means and standard deviations were calculated from at least 100 punctual signals in representative areas of rosette cells from five independent lines for each construct.

Homologs of NAIPs in Other Organisms

To identify homologs of NAIP genes in other organisms, we searched GenBank's nonredundant protein database. Homologs of NAIP genes are not found in the archaea, eubacteria, or fungi. No NAIP genes are identified in the linkage of animals either, except in Chinese rufous horseshoe bat (*Rhinolophus sinicus*), whose sequenced genome contains three NAIP homologs (GenBank accessions XP_019577945.1, XP_019578097.1, and XP_019578712.1). Interestingly, the three bat NAIP homologs are more than 90% identical in both amino acid and DNA sequences with the NAIP1, NAIP 2, and NAIP 3, respectively, from Arabidopsis and their homologs in the Brassicaceae. These levels of sequence homology of the NAIP homologs between the bat and the Brassicaceae species are even higher than those between different plant families in the Brassicales. No NAIP homologs are found in the genomes of other sequenced bats or mammals. These observations indicate that the NAIP genes found in Chinese rufous horseshoe bat likely resulted

from sequencing contamination or recent horizontal gene transfer. Besides the three NAIP genes from the bat, we identified more than 20 NAIP genes from nonplant eukaryotic species. All these species are in the kingdom of Protista, most belonging to the phylum of Apicomplexa in the large clade of parasitic alveolate. These species of protists, all with a single NAIP homolog, include those in the genera of *Besnoitia*, *Cystoisospora*, *Plasmodium*, *Toxoplasma*, *Eimeria*, *Chrysochromulina*, *Thecamonas*, *Theileria*, *Babesia*, and *Perkinsus*. Amino acid sequence analysis revealed that the NAIP proteins from protists all contain the conserved N-terminal CC and C-terminal HHD domains with a variable central domain (Supplemental Fig. S10). Therefore, the domain architecture of the protist NAIP homologs is identical to those of Arabidopsis NAIP proteins. However, the middle domains of these protist NAIP proteins contain no or few SP or TP phosphorylation motifs (Supplemental Fig. S10).

In the linkage of plants, we identified a single NAIP homolog in *Chlamydomonas reinhardtii*, a unicellular green alga with its evolutionary position located before the divergence of land plants (Manhart, 1994; An et al., 1999). Other sequenced chlorophytes including *Dunaliella salina*, *Volvox carteri*, *Coccomyxa subellipsodea*, and *Micromonas pusilla* also contain a single NAIP homolog in their sequenced genomes. There are six NAIP homologs in the moss *Physcomitrella patens*, an early diverged land plant. However, two other embryophytes, the moss species *Sphagnum fallax* and liverwort *Marchantia polymorpha*, each contain only four NAIP homologs. There are three NAIP homologs in the fern *Selaginella moellendorffii*, an ancient vascular plant (Banks et al., 2011). In angiosperms, rice (*Oryza sativa*) contains four NAIP homologs, whereas both tomato (*Solanum lycopersicum*) and *Medicago* contain three NAIP homologs, as does Arabidopsis. In gymnosperms, genomes from *Picea abies*, *Picea glauca*, and *Pinus taeda* have been sequenced but not well annotated (Birol et al., 2013; Nystedt et al., 2013; Neale et al., 2014; Zimin et al., 2014). A preliminary search revealed that these sequenced genomes each have three to four NAIP homologs as well. Thus, NAIP proteins have originated in early eukaryotes and are present in all branches of land plants as a small family with three to four members. We compared the amino acid sequences of these NAIP proteins from plant species using multisequence alignment and confirmed that the C-terminal HHD domains are the most conserved regions among these NAIP homologs (Supplemental Fig. S11). Substantial conservation was also observed among the N-terminal CC domains of the NAIP homologs, but their middle domains are more divergent (Supplemental Fig. S11).

To examine the evolutionary relationship among the NAIP proteins, we generated a phylogenetic tree with 27 NAIP protein sequences from three protists (*Plasmodium retictum*, *Cryptosporidium suis*, and *Eimeria necatrix*), *C. reinhardtii*, *P. patens*, *S. moellendorffii*, Arabidopsis, tomato, *Medicago*, and rice. As shown in Figure 9, we observed only six branches with bootstrap values larger than 90%. A majority of the branches, particularly the deep ones, had low confidence values, likely due to the divergent sequences of the protein family. When only those homologs from the plant linkage were included, the generated phylogenetic tree also contains only a few branches with larger than 90% bootstrap values (Supplemental Figure S12). Almost all those branches with larger than 90% bootstrap values are between NAIP homologs from the same species (Fig. 9; Supplemental Fig. S12). As a result, even though

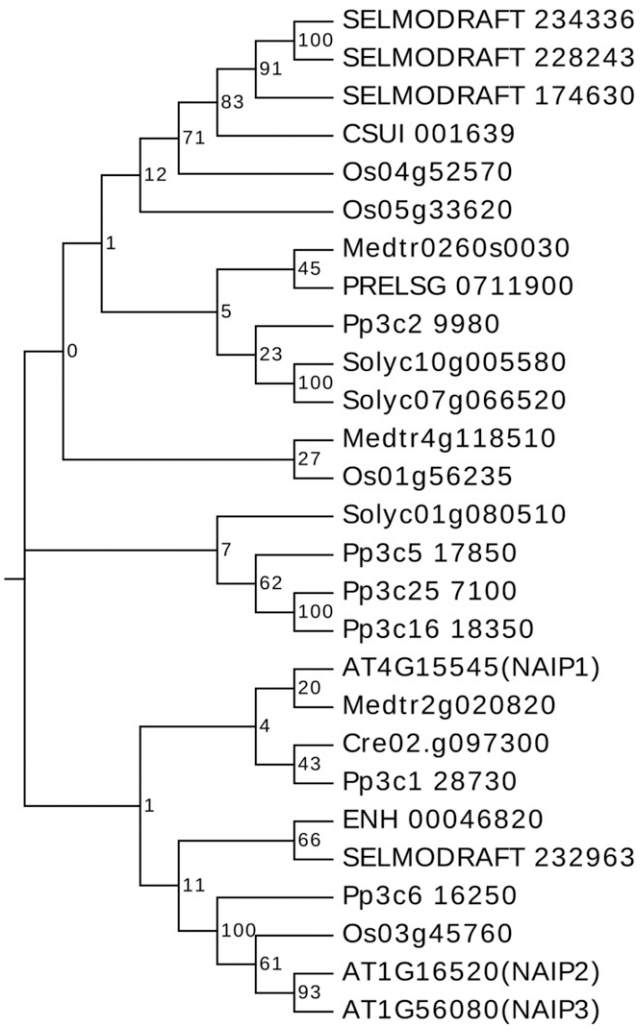

Figure 9. The phylogenetic relationship of NAIP homologs from plants and protists. The tree was inferred using the neighbor-joining method. Phylogenetic analyses were conducted in MEGA5. Bootstrap values from 1000 replicates were used to assess the robustness of the tree. NAIP homologs in the phylogenetic analysis include those from *P. retictum* (PRELSG0711900), *C. suis* (CSUI001639), *E. necatrix* (ENH00046820), *C. reinhardtii* (Cre02.g097300), *P. patens* (Pp3c16_18350V3, Pp3c5_17850V3, Pp3c25_7100V3, Pp3c6_16250V3, Pp3c1_28730V3, Pp3c2_9980V3, and Pp3c14_14570V3), *S. moellendorffii* (SELMODRAFT234336, SELMODRAFT 228243, SELMODRAFT174630, and SELMODRAFT232963), Arabidopsis (At4g15545, At1g16520, and At1g56080), tomato (Solyc01g080510, Solyc10g005580, and Solyc07g066520), *Medicago* (Medtr4g118510, Medtr2g020820, and Medtr0260s0030), and rice (Os05g33620, Os03g45760, Os04g52570, and Os01g56235).

most of these species, particularly the land plants, have a similar number of NAIP homologs, they cannot be grouped into distinct subfamilies with high confidence (Fig. 9; Supplemental Fig. S12). A clear evolutionary relationship among the NAIP homologs from low eukaryotes to angiosperm plants was not apparent either (Fig. 9; Supplemental Fig. S12). On the other hand, it is apparent from the phylogenetic trees that Arabidopsis NAIP2 and NAIP3 are structurally more closely related to each than to NAIP1 (Fig. 9; Supplemental Fig. S12), consistent with their functional relationship in the association with different types of ER-derived compartments.

DISCUSSION

The ER bodies found only in the Brassicales order contain an abundant level of PYK10 myrosinase proteins, which hydrolyze indole glucosinolates to chemically reactive products toxic to pathogens and herbivores (Nakano et al., 2017). Among those components identified from the studies of the ER body, NAI1, NAI2, and PYK10 play roles in the formation of the ER-derived compartment based on their mutant phenotypes (Nakano et al., 2014). As a transcription factor, NAI1 is important for the ER body formation by regulating expression of important genes, including PKY10 and NAI2 associated with the ER bodies (Matsushima et al., 2004). PYK10 is the major cargo of the ER body but also affects the formation of the ER body based on altered shape of the ER bodies in its mutant (Nagano et al., 2009). NAI2, on the other hand, is an essential ER body component and is present only in the Brassicales order just like the ER body (Yamada et al., 2008), making it a uniquely important factor for understanding the biogenesis and evolution of the ER body. Arabidopsis TSA1 is a close homolog of NAI2 and plays a critical role in jasmonic-acid-induced ER body formation (Geem et al., 2019). Using yeast two-hybrid screens, we identified a small NAIP protein family whose three members all interact with NAI2 (Fig. 1). To determine the functional relevance of the interactions of the three NAIP proteins with NAI2, we analyzed their roles using a genetic approach. Formation of the ER bodies was greatly reduced in the *naip1/naip2/naip3* triple mutants, as in the *nai2* mutants (Fig. 7). These results indicate the three NAIP proteins function together with NAI2 as critical ER body components with an important role in the formation of the ER-derived compartments. However, NAIP proteins differ from NAI2 in two important ways. First, while NAI2 is found only in the Brassicales (Yamada et al., 2008), homologs of the NAIP proteins are identified not only in photosynthetic organisms, but also in some protists (Fig. 9; Supplemental Fig. S12). Second, while NAI2 is associated only with the ER bodies normally present abundantly in the cotyledons, hypocotyls, and roots, but not in rosette leaves, two members of the NAIP protein family, NAIP2 and NAIP3, are present both in the ER bodies and in other types of vesicular structures that are also likely to be derived from the ER and present not only in the cotyledons, hypocotyls, and roots but also in the rosette leaves (Fig. 5). Based on these findings, we propose that the NAI2-containing ER bodies in the Brassicales may have evolved from NAIP-containing ER-derived structures widely present not only in plants but also in protists.

The proposed evolution of the NAI2-containing ER bodies from NAIP-containing ER-derived structures is consistent with the functional diversification of the NAIP protein family in Arabidopsis. Confocal fluorescence microscopy analysis of GFP fusion proteins revealed that NAIP1-containing vesicles, just like the PYK10- and NAI2-containing ER bodies, are found abundantly only in the cotyledons, hypocotyls, and roots, but not in the rosette leaves of mature plants (Fig. 5). In addition, formation of the NAIP-containing ER-derived structures is dependent on NAI2 (Fig. 5). NAIP2 and NAIP3 also form the ER bodies based on their colocalization with PYK10 in the cotyledon cells (Fig. 4) and play an important redundant role with NAIP1 in the ER body formation (Fig. 7). However, NAIP2- and NAIP3-containing structures are found not only in the cotyledons, hypocotyls, and roots, but also in the rosette leaves of mature plants (Fig. 5). Furthermore, NAIP2- and NAIP3-containing structures in these tissues are still observed in the *nai2* mutant plants (Fig. 7). Thus, while NAIP1 has evolved to function specifically for ER body formation, NAIP2 and NAIP3 are less specialized and can function as components of not only the ER bodies but also other ER-derived structures that can be formed in a wider range of plant tissues. The functional diversification of the NAIP proteins supports that NAIP-containing vesicular structures, which are likely present widely in photosynthetic eukaryotes and some protists, have evolved into different subtypes, including the ER bodies in plants in the Brassicales order.

Besides their tissue specificity and NAI2 dependency, the NAIP1-labeled ER bodies and NAIP3-labeled ER compartments differ in additional characteristics. For example, the NAIP1-labeled ER bodies are closely connected with and continuous to the ER (Fig. 5). The mechanism for the maintenance of such a continuous network between the ER bodies and the ER is unknown but could result from the budding of the ER bodies, on the one hand, and the retrograde incorporation of the ER bodies back into the ER. On the other hand, a majority of NAIP3-labeled vesicles appear to be well separated from the ER (Fig. 5). In addition, NAIP1-labeled ER bodies have similar sizes of around 1 μm in diameter (Fig. 6). However, NAPI3-labeled structures are heterogeneous in size, with their diameters ranging from 0.5 to 5 μm (Fig. 6). One possible reason for the wide size range of NAIP3-containing structures is the presence of distinct types of ER-derived structures with different cargo molecules. To determine the structural basis for the distinct characteristics of these ER-derived compartments, we performed domain swapping and generated

chimeric NAIP proteins from NAIP1 and NAIP3. By expressing their GFP fusion genes in the transgenic plants, we compared the ability to form punctual signals in the cotyledons and rosette leaves in the transgenic plants, from which a critical role of the N-terminal CC domain in the functional differentiation between NAIP1 and NAIP3 started to emerge. First, as with the ER bodies, NAIP1-containing compartments are formed abundantly only in the cotyledon, but not in the rosette leaves under normal conditions (Fig. 8). Replacing the middle domain and the C-terminal HHD domain of NAIP1 with the corresponding domains of NAIP3 did not change the tissue specificity of the ER-derived compartments (Fig. 8). However, replacing the N-terminal domain of NAIP1 with the corresponding domain of NAIP3 led to formation of punctual signals not only in the cotyledons, but also in the rosette leaves (Fig. 8). Second, in both the cotyledons and rosette leaves, those chimeric NAIP proteins containing the NAIP1 N-terminal domain formed punctual signals with sizes similar to those of NAIP1 (Fig. 8). On the other hand, those chimeric NAIP proteins containing the NAIP3 N-terminal domain form punctual signals with a wide range of sizes, as found with the NAIP3-containing structures (Fig. 8). Thus, the N-terminal domains of the NAIP proteins play a major role in determining the tissue specificity and the physical size and shape of the ER-derived structures. The N-terminal CC domains of these NAIP proteins are involved in self-interactions and mutual interactions (Fig. 1), and it is unclear how these interacting domains share the same or similar specificity of protein interactions but have differential ability to regulate the tissue specificity and physical size of the ER-derived structures. One possibility could be that these N-terminal CC domains, besides their self-interactions and mutual interactions, also interact with other proteins with different specificities. In addition, NAIP1 contains a highly acidic motif (EIEEEEEE) at its N terminus, which is absent in NAIP2 and NAIP3 (Supplemental Fig. S1). A similar highly acidic motif is also present in the NAIP1 homologs from the Brassicales species. NAIP homologs from plants outside of the Brassicales order generally do not contain such a highly acidic motif at the N terminus (Supplemental Fig. S11). However, among the four rice NAIP homologs, one contains such a highly acidic motif at its N terminus (Supplemental Fig. S11). Further structural dissection will be necessary to determine whether the addition of the acidic motif is a critical structural determinant for the functional diversification of NAIP proteins for association with special ER-derived structures.

Recent studies have strongly suggested that the ER bodies have coevolved with the indole glucosinolate metabolism as a mechanism of plant defense against pathogens and herbivores (Nakano et al., 2017). The finding that there are likely ER-derived structures similar to the NAIP2- and NAIP3-containing structures in all plants raises questions about their biological functions. ER-derived compartments in plants often contain storage proteins and lipid molecules for transport to storage vacuoles in seeds and other storage tissues. As expression of NAIP genes is mostly in nonstorage tissues, they are unlikely to be associated with storage protein or lipid bodies. The *naip1/naip2/naip3* triple mutant plants are indistinguishable from wild-type plants in growth and development, suggesting that NAIP proteins probably play roles in plant responses to environmental conditions. Given the involvement of NAIP1-labeled ER bodies in indole glucosinolate metabolism, NAIP2- and NAIP3-labeled compartments could contain enzymes in the biosynthesis and metabolism of other metabolites during plant responses to environmental stimuli. Interestingly, even though the NAIP proteins are found in all plants, their amino acid sequences have relatively low homology among different plants (Fig. 9; Supplemental Fig. S11), which may lead to the formation of different ER-derived compartments with diverse cargo. In Arabidopsis, the size of NAIP3-labeled ER structures is highly heterogeneous, suggesting that the same protein may be associated with different types of ER-derived structures that accumulate or transport different types of cargo molecules with distinct biological functions. To address these possibilities and establish the biological functions of these ER-derived compartments, it will be necessary to isolate these vesicles, identify the cargo, determine their biochemical and molecular activities, and establish the associated biological processes. With the identification and characterization of these NAIP proteins, it is now possible to develop experimental approaches to address these important questions.

MATERIALS AND METHODS

Plant Growth Conditions and Arabidopsis Genotypes

Arabidopsis (*Arabidopsis thaliana*) plants were grown in growth chambers or growth rooms at 23°C with 120 $\mu E \cdot m^{-2} \cdot s^{-1}$ light intensity and a photoperiod of 12 h of light and 12 h of dark. The *ubac2a-1* and *ubac2b-1* mutants have been previously described (Zhou et al., 2018). Homozygous *nai2-2* (Salk_005896), *naip1-1* (FLAG_512D04), *naip2-1* (GABI_529B11), *naip2-2* (GABI_922B07), and *naip3-1* (Salk_200721) were identified by PCR using gene-specific primers flanking the T-DNA/transposon insertions (Supplemental Table S1).

Yeast Two-Hybrid Screen and Assays

In order to find UBAC2a-interacting proteins, a Gal4-based yeast two-hybrid system was utilized as previously described (Zhou J et al., 2018). In brief, UBAC2a was PCR amplified using gene-specific primers (5′-agcgaattcatgaacggcggtccctcc-3′ and 5′-agcgtcgacttagtttctgtcgaatcccatt-3′) and cloned into pBD-GAL4 vector to generate the bait vector. The Arabidopsis HybridZAP-2.1 two-hybrid cDNA library was prepared from Arabidopsis plants as previously described (Xu et al., 2006). The bait plasmid and the cDNA library were used to transform yeast strain YRG-2. Yeast transformants were plated onto selection medium lacking Trp, Leu, and His and confirmed by β-galactosidase activity assays using *o*-nitrophenyl-β-D-galactopyranose as substrate.

To identify NAIP1-interacting proteins, full-length NAIP1 was PCR amplified using gene-specific primers (5′-agcgagctcatgtcagagatagaagaagaagaaga-3′ and 5′-agcggatccttagtgagcgttgcgtgtg-3′) and cloned into pBD-GAL4 vector as bait vector in screening the Arabidopsis HybridZAP-2.1 two-hybrid cDNA library for interacting proteins. For assays of interactions among NAIP2 and NAIP3 with NAI2 and NAIP1, full-length NAIP2 was PCR amplified using gene-specific primers (NAIP2, 5′-agcactagtatgggagacgaccagctaga-3′ and 5′-agc

ggatccttaacgcatgtttctgttaagg-3′; NAIP3, 5′-agcgagctcatgtctcaaagcggcggt-3′ and 5′-agcactagttcaacgcccacttgtgagaagtc-3′) and cloned into the bait vector pBD-GAL4. The DNA fragment for the C-terminal HHD domain of NAIP1 was PCR amplified using the primers 5′-agcggatccttagtgagcgttgcgtgtg-3′ and 5′-agc gaattctctcctcgacgccattctgt-3′, and the PCR product was inserted into the pBD-GAL4 vector. The DNA fragment for the N-terminal domain of NAIP1 was generated by removing the C-terminal *Bgl*II-*Bam*HI fragment from the NAIP1 coding sequence in the pBD-NAIP1 vector. Various combinations of bait and prey constructs were cotransformed into yeast cells, and interactions were analyzed through assays of the LacZ β-galactosidase reporter gene activity.

BiFC Assay

The BiFC vectors pFGC-N-YFP and pFGC-C-YFP have been previously described (Cui et al., 2007). The full-length NAIP1, NAIP2, NAIP3, NAI2, and *PYK10* coding sequences were PCR amplified using gene-specific primers (NAIP1, 5′-agcgagctcatgtcagagatagaagaagaagaagaa-3′ and 5′-agcggatccgtgagc gttgcgtgtgatca-3′; NAIP2, 5′-agcggatccatgggagacgaccagctaga-3′ and 5′-agctct agaacgcatgtttctgttaaggag-3′; NAIP3, 5′-agcgagctcatgtctcaaagcggcggt-3′ and 5′-agctctagaacgcccacttgtgagaagtc-3′; NAI2, 5′-agcgagctcatgggaacaaagtttttagctct-3′ and 5′-agcggatccattaagtgaactaagaaactcaaccca-3′; PYK10, 5′-agcactagtatggtt ttgcaaaagcttcc-3′ and 5′-agcggatccaagctcatccttcttgagc-3′) and cloned into pFGC-N-YFP or pFGC-C-YFP. The plasmids were introduced into *Agrobacterium tumefaciens* (strain GV3101) and transformed into Arabidopsis plants as described previously (Clough and Bent, 1998). Transgenic plants were identified based on the Basta resistance. Positive T1 transformants were crossed to generate transgenic plants containing a pair of BiFC constructs. BiFC fluorescent signals were analyzed in the Arabidopsis cotyledon by a Zeiss LSM710 confocal microscope with appropriate filter sets (excitation 514 nm, emission 525–555 nm). The images were superimposed with Zeiss LSM710 software.

Assays of GUS Activity and Histochemical Staining

The ~1.5-kb promoter regions of the NAIP genes were PCR amplified from the genomic DNA using gene-specific primers (NAIP1, 5′-agcggatccttttggatttgt tggtgacg-3′ and 5′-agcactagtttctcgttttctgtgttttttttg-3′; NAIP2, 5′-agcggatcccgc tgtcctacgaacgtacc-3′ and 5′-agcactagtttctccaccacaagaatgtcc-3′; and NAIP3, 5′-agcggatccgttcaaaccaaatggcctgt-3′ and 5′-agcccatggaaagggccctcgggatct-3′). The promoter sequences were cloned into a modified pFGC5941 binary vector that contains a promoterless *GUS* gene. The promoter-GUS constructs were transformed into Arabidopsis plants using the *Agrobacterium*-mediated floral-dip method (Clough and Bent, 1998). Transgenic plants were obtained based on the Basta resistance. At least five independent lines were selected for the histochemical staining of GUS activity as described previously (Jefferson et al., 1987). In brief, 1-week-old seedlings, 3-week-old rosette leaves, and the inflorescence of the transgenic plants were prefixed in ice-cold 90% acetone for 20 min and then immersed in staining solution (50 mM sodium phosphate buffer, pH 7.0, 10 mM EDTA, 2 mM potassium ferricyanide, 2 mM potassium ferrocyanide, 0.1% [v/v] Triton X-100, and 2 mM 5-bromo-4-chloro-3-indolyl-β-D-GlcA) at 37°C for 16 h. Tissues were rinsed in 70% (v/v) ethanol, and the solution was changed every hour until the chlorophyll was removed. Micrographs were taken using a Nikon Digital Sight DS-5M camera attached to a Nikon Eclipse 80i compound microscope.

Plasmid Construction and Generation of Transgenic Arabidopsis Plants

For generating NAIP-GFP fusion genes, full-length NAIP coding sequences were PCR amplified using gene-specific primers (*NAIP1*, 5′-agcgagctcatgtcagag atagaagaagaagaaga-3′ and 5′-agcggatccgtgagcgttgcgtgtgatc-3′; *NAIP2*, 5′-agc gtcgacatgggagacgaccagctaga-3′ and 5′-agcggatccacgcatgtttctgttaaggag-3′; *NAIP3*, 5′-agcgagctcatgtctcaaagcggcggt-3′ and 5′-agcactagtacgcccacttgtgagaag tcc-3′), fused to the GFP gene behind the *CaMV 35S* promoter in a modified pCAMBIA1300 plant transformation vector, and transformed into Arabidopsis wild-type and *nai2* mutant plants. The cotyledons, hypocotyls, and roots of 1-week-old transgenic seedlings and rosette leaves of 3- to 4-week-old transgenic Arabidopsis plants were used for imaging of GFP using standard confocal laser microscopy. At least 10 independent lines for each construct were used in these experiments.

For generating mCherry-labeled PYK10 and NAI2 ER body markers, full-length PYK10 and NAI2 coding sequences were PCR amplified using gene-specific primers (PYK10, 5′-agcgaattcatggttttgcaaaagcttcc-3′ and 5′-agcggatcc aaagctcatccttcttgagc-3′; NAI2, 5′-agcgaattcatgggaacaaagtttttagctctg-3′ and 5′-agcggatcccattaagtgaactaagaaactcaacccaaa-3′), fused to an mCherry gene behind the *CaMV 35S* promoter in a modified pCAMBIA1300 plant transformation vector, and transformed into Arabidopsis plants. The cotyledon, hypocotyl, and root of 1-week-old transgenic seedlings and rosette leaves of 3- to 4-week-old stably transformed Arabidopsis plants from at least 10 independent lines for each construct were used for imaging of mCherry using standard confocal laser microscopy.

To generate chimeric constructs for NAIP proteins, the three domains between NAIP1 and NAIP3 were swapped using overlapping PCR. The N-terminal, middle, and C-terminal domains of NAIP1 are defined as amino acid residues 1 to 141, 142 to 270, and 271 to 337, respectively. The N-terminal, middle, and C-terminal domains of NAIP3 are defined as amino acid residues 1 to 127, 128 to 243, and 244 to 310, respectively. The chimeric genes were fused to the *GFP* gene behind the *CaMV 35S* promoter in a modified pCAMBIA1300 plant transformation vector and transformed into Arabidopsis plants. Transgenic lines were then selected based on the hygromycin resistance. At least six lines for each construct were chosen for examination of GFP signals.

Subcellular Localization

For subcellular colocalization in Arabidopsis, transgenic NAIP-GFP plants were crossed with transgenic PYK10-mCherry lines, and the F1 progeny was used for the subcellular colocalization. Imaging of coexpressed GFP and mCherry signals was performed with standard confocal laser microscopy with appropriate filter sets: GFP (excitation 488 nm, emission 500–550 nm) and mCherry (excitation 561 nm, emission 570–620 nm).

Phylogenetic Analysis of NAIP Proteins

The full-length protein sequences of NAIP homologs were retrieved from the GenBank and Phytozome databases. The phylogenetic trees were constructed using the neighbor-joining method with MEGA5.

Protease Protection Assays and Western Blotting

For generating NAI2- and PYK10-myc fusion genes, full-length *NAI2* and *PYK10* coding sequences were PCR amplified using gene-specific primers (*NAI2*, 5′-agcccatgggaacaaagtttttagctct-3′ and 5′-agcttaattaaattaagtgaactaaga aactcaaccca-3′; *PYK10*, 5′-agcgtcgacatggttttgcaaaagcttcc-3′ and 5′-agcttaattaaa agctcatccttcttgagc-3′), fused to an 4xmyc tag sequence behind the *CaMV 35S* promoter in a modified pFGC5941 plant transformation vector, and transformed into Arabidopsis plants. Microsomal fractions were isolated from the leaves of the transgenic plants by differential centrifugation as previously described (Abas and Luschnig, 2010). Membrane pellets containing the myc-tagged NAI2 or PYK10 proteins were resuspended in an assay buffer (20 mM Tris/HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, and 10% glycerol) and split into two samples. Triton X-100 was added to one sample to a final concentration of 0.1%. After incubation on ice for 10 min, each sample was further split into five microcentrifuge tubes and subjected to 2 μg/mL proteinase K digestion in the absence or presence of 0.1% of Triton X-100 for 0 to 40 min on ice. The reactions were stopped by adding 4 mM phenylmethanesulfonyl fluoride. The samples were incubated at 100°C for 15 min in the Laemmli sample buffer and separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The myc-tagged NAI2 and PYK10 proteins were detected by western blotting using an anti-myc monoclonal antibody as previously described (Zhou et al., 2018).

Accession Numbers

Arabidopsis Genome Initiative numbers for the genes discussed in this article are as follows: NAI1 (At2g22770), NAI2 (At3g15950), PYK10 (At3g09260), NAIP1 (At4g15545), NAIP2 (At1g16520), NAIP3 (At1g56080), TSA1 (At1g52410), MEB1 (At4g27860), and MEB2 (At5g24290).

Supplemental Data

The following supplemental materials are available:

Supplemental Figure S1. Sequences and structures of Arabidopsis NAIP proteins.

Supplemental Figure S2. Western blot analysis of transgenic Arabidopsis plants containing different combinations of BiFC constructs.

Supplemental Figure S3. BiFC assays of NAIP-NAI2 interactions in Arabidopsis cotyledons.

Supplemental Figure S4. Colocalization of BIFC and PYK1-mCherry fluorescent signals.

BiFC assays of NAIP-NAI2 interactions in Arabidopsis cotyledons.

Supplemental Figure S5. Protease K protection assays of myc-tagged NAI2 and PYK10.

Supplemental Figure S6. Role of NAI2 in the formation of punctate structures containing NAIP proteins.

Supplemental Figure S7. T-DNA insertion mutants for the NAIP genes.

Supplemental Figure S8. ER body formation in the *naip1/naip2* and *ubac2a/2b* double mutants.

Supplemental Figure S9. Complementation of *naip1/ naip2/ naip3* triple mutant by the GFP fusions of NAIP proteins.

Supplemental Figure S10. Amino acid sequence alignment of Arabidopsis NAIP proteins and the NAIP homologs from protists.

Supplemental Figure S11. Amino acid sequence alignment of NAIP homologs from the plant linkage.

Supplemental Figure S12. The phylogenetic relationship of NAIP homologs from plants. The tree was inferred using the neighbor-joining method.

Supplemental Table S1. PCR primers for *naip* and *nai2-2* mutant screens.

ACKNOWLEDGMENTS

We thank the Arabidopsis Resource Center at the Ohio State University for the Arabidopsis mutants.

Received December 3, 2018; accepted February 5, 2019; published February 15, 2019.

LITERATURE CITED

Abas L, Luschnig C (2010) Maximum yields of microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation. Anal Biochem **401:** 217–227

An SS, Möpps B, Weber K, Bhattacharya D (1999) The origin and evolution of green algal and plant actins. Mol Biol Evol **16:** 275–285

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, et al (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science **332:** 960–963

Benham AM (2012) Protein secretion and the endoplasmic reticulum. Cold Spring Harb Perspect Biol **4:** a012872

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP, et al (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics **29:** 1492–1497

Chrispeels MJ, Herman EM (2000) Endoplasmic reticulum-derived compartments function in storage and as mediators of vacuolar remodeling via a new type of organelle, precursor protease vesicles. Plant Physiol **123:** 1227–1234

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of Arabidopsis thaliana. Plant J **16:** 735–743

Cui X, Fan B, Scholz J, Chen Z (2007) Roles of Arabidopsis cyclin-dependent kinase C complexes in cauliflower mosaic virus infection, plant growth, and development. Plant Cell **19:** 1388–1402

Geem KR, Kim DH, Lee DW, Kwon Y, Lee J, Kim JH, Hwang I (2019) Jasmonic acid-inducible TSA1 facilitates ER body formation. Plant J **97:** 267–280.

Hara-Nishimura I, Matsushima R, Shimada T, Nishimura M (2004) Diversity and formation of endoplasmic reticulum-derived compartments in plants. Are these compartments specific to plant cells? Plant Physiol **136:** 3435–3439

Hawes C, Saint-Jore C, Martin B, Zheng HQ (2001) ER confirmed as the location of mystery organelles in Arabidopsis plants expressing GFP! Trends Plant Sci **6:** 245–246

Hayashi Y, Yamada K, Shimada T, Matsushima R, Nishizawa NK, Nishimura M, Hara-Nishimura I (2001) A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. Plant Cell Physiol **42:** 894–899

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6:** 3901–3907

Kumar MN, Hsieh YF, Verslues PE (2015) At14a-Like1 participates in membrane-associated mechanisms promoting growth during drought in *Arabidopsis thaliana*. Proc Natl Acad Sci USA **112:** 10545–10550

Lee SY, Voronov S, Letinic K, Nairn AC, Di Paolo G, De Camilli P (2005) Regulation of the interaction between PIPKI gamma and talin by proline-directed protein kinases. J Cell Biol **168:** 789–799

Li W, Zang B, Liu C, Lu L, Wei N, Cao K, Deng XW, Wang X (2011) TSA1 interacts with CSN1/CSN and may be functionally involved in Arabidopsis seedling development in darkness. J Genet Genomics **38:** 539–546

Manhart JR (1994) Phylogenetic analysis of green plant rbcL sequences. Mol Phylogenet Evol **3:** 114–127

Matsushima R, Hayashi Y, Kondo M, Shimada T, Nishimura M, Hara-Nishimura I (2002) An endoplasmic reticulum-derived structure that is induced under stress conditions in Arabidopsis. Plant Physiol **130:** 1807–1814

Matsushima R, Kondo M, Nishimura M, Hara-Nishimura I (2003) A novel ER-derived compartment, the ER body, selectively accumulates a beta-glucosidase with an ER-retention signal in *Arabidopsis*. Plant J **33:** 493–502

Matsushima R, Fukao Y, Nishimura M, Hara-Nishimura I (2004) NAI1 gene encodes a basic-helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. Plant Cell **16:** 1536–1549

Nagano AJ, Maekawa A, Nakano RT, Miyahara M, Higaki T, Kutsuna N, Hasezawa S, Hara-Nishimura I (2009) Quantitative analysis of ER body morphology in an *Arabidopsis* mutant. Plant Cell Physiol **50:** 2015–2022

Nakano RT, Yamada K, Bednarek P, Nishimura M, Hara-Nishimura I (2014) ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. Front Plant Sci **5:** 73

Nakano RT, Piślewska-Bednarek M, Yamada K, Edger PP, Miyahara M, Kondo M, Böttcher C, Mori M, Nishimura M, Schulze-Lefert P, et al (2017) PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum bodies and glucosinolates in *Arabidopsis thaliana*. Plant J **89:** 204–220

Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, et al (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol **15:** R59

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature **497:** 579–584

Ogasawara K, Yamada K, Christeller JT, Kondo M, Hatsugai N, Hara-Nishimura I, Nishimura M (2009) Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct beta-glucosidases. Plant Cell Physiol **50:** 480–488

Sherameti I, Venus Y, Drzewiecki C, Tripathi S, Dan VM, Nitz I, Varma A, Grundler FM, Oelmüller R (2008) PYK10, a beta-glucosidase located in the endoplasmatic reticulum, is crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus *Piriformospora indica*. Plant J **54:** 428–439

Suzuki T, Nakajima S, Morikami A, Nakamura K (2005) An *Arabidopsis* protein with a novel calcium-binding repeat sequence interacts with TONSOKU/MGOUN3/BRUSHY1 involved in meristem maintenance. Plant Cell Physiol **46:** 1452–1461

Viotti C, Krüger F, Krebs M, Neubert C, Fink F, Lupanga U, Scheuring D, Boutté Y, Frescatada-Rosa M, Wolfenstetter S, et al (2013) The endoplasmic reticulum is the main membrane source for biogenesis of the lytic vacuole in Arabidopsis. Plant Cell **25:** 3434–3449

Vitale A, Denecke J (1999) The endoplasmic reticulum-gateway of the secretory pathway. Plant Cell **11:** 615–628

Wang JZ, Li B, Xiao Y, Ni Y, Ke H, Yang P, de Souza A, Bjornson M, He X, Shen Z, et al (2017) Initiation of ER body formation and indole glucosinolate metabolism by the plastidial retrograde signaling metabolite, MEcPP. Mol Plant **10:** 1400–1416

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell **18:** 1310–1326

Yamada K, Nagano AJ, Nishina M, Hara-Nishimura I, Nishimura M (2008) NAI2 is an endoplasmic reticulum body component that enables ER body formation in *Arabidopsis thaliana*. Plant Cell **20:** 2529–2540

Yamada K, Nagano AJ, Nishina M, Hara-Nishimura I, Nishimura M (2013) Identification of two novel endoplasmic reticulum body-specific integral membrane proteins. Plant Physiol **161:** 108–120

Zhou J, Wang Z, Wang X, Li X, Zhang Z, Fan B, Zhu C, Chen Z (2018) Dicot-specific ATG8-interacting ATI3 proteins interact with conserved UBAC2 proteins and play critical roles in plant stress responses. Autophagy **14:** 487–504

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ, et al (2014) Sequencing and assembly of the 22-gb loblolly pine genome. Genetics **196:** 875–890