The Ubiquitous Distribution of Late Embryogenesis Abundant Proteins across Cell Compartments in Arabidopsis Offers Tailored Protection against Abiotic Stress

Adrien Candat, a,b Gaël Paszkiewicz, a Martine Neveu, b Romain Gautier, c David C. Logan, a Marie-Hélène Avelange-Macherel, d and David Macherel a,1

a Université d’Angers, UMR 1345 Institut de Recherche en Horticulture et Semences, F-49045 Angers, France
b INRA, UMR 1345 Institut de Recherche en Horticulture et Semences, F-49045 Angers, France
c Institut de Pharmacologie Moléculaire et Cellulaire, Université de Nice Sophia-Antipolis and Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7275, F-06560 Valbonne, France
d Agrocampus Ouest, UMR 1345 Institut de Recherche en Horticulture et Semences, F-49045 Angers, France

ORCID IDs: 0000-0002-8980-240X (D.C.L.); 0000-0001-6002-3130 (M.-H.A.-M.); 0000-0002-3352-2185 (D.M.)

Late embryogenesis abundant (LEA) proteins are hydrophilic, mostly intrinsically disordered proteins, which play major roles in desiccation tolerance. In Arabidopsis thaliana, 51 genes encoding LEA proteins clustered into nine families have been inventoried. To increase our understanding of the yet enigmatic functions of these gene families, we report the subcellular location of each protein. Experimental data highlight the limits of in silico predictions for analysis of subcellular localization. Thirty-six LEA proteins localized to the cytosol, with most being able to diffuse into the nucleus. Three proteins were exclusively localized in plastids or mitochondria, while two others were found dually targeted to these organelles. Targeting cleavage sites could be determined for five of these proteins. Three proteins were found to be endoplasmic reticulum (ER) residents, two were vacuolar, and two were secreted. A single protein was identified in pexophagosomes. While most LEA protein families have a unique subcellular localization, members of the LEA_4 family are widely distributed (cytosol, mitochondria, plastid, ER, and pexophagosome) but share the presence of the class A α-helix motif. They are thus expected to establish interactions with various cellular membranes under stress conditions. The broad subcellular distribution of LEA proteins highlights the requirement for each cellular compartment to be provided with protective mechanisms to cope with desiccation or cold stress.

INTRODUCTION

During late development, before they enter the desiccation phase, plant seeds accumulate a class of polypeptides known as LEA (late embryogenesis abundant) proteins, which share common properties such as low sequence complexity, repeat motifs, high hydrophilicity, and often a lack of ordered structure in the native state (Tunnacliffe and Wise, 2007; Shih et al., 2008; Tunnacliffe et al., 2010). Although there is a great diversity in LEA proteins, most can be grouped into eight families in the PFAM database (Finn et al., 2010) according to their primary sequences: dehydrin, LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_6, and seed maturation protein (SMP). The pertinence of such classification was reinforced and extended by computational analysis of the physico-chemical properties of more than 700 LEA protein sequences deposited in LEAPdb, a publicly available database (Hunault and Jaspard, 2010), which allowed a more detailed organization of the proteins into 12 non-redundant classes (Jaspard et al., 2012). Interestingly, LEA proteins have also been discovered in various species or phyla of invertebrates (bdelloid rotifers, nematodes, tardigrades, and arthropods), which are all anhydrobiotes (Tunnacliffe and Wise, 2007). In these organisms, as in the case of plant seeds, LEA proteins accumulate prior to desiccation, which is a very strong argument for a role of these proteins in desiccation tolerance.

LEA genes are highly represented in plant genomes with, for instance, 51 LEA genes in Arabidopsis thaliana (Bies-Ethève et al., 2008; Hundertmark and Hincha, 2008). According to these studies, 47 of the corresponding LEA proteins are distributed within the eight PFAM families. Two proteins were considered to be members of the LEA_4 family by Hundertmark and Hincha (2008), who also included two other proteins without PFAM classification in a group which we will also refer to as AtM.

Quantitative RT-PCR analysis showed that most Arabidopsis LEA genes were highly expressed in seeds (Hundertmark and Hincha, 2008), in agreement with the accumulation of LEA proteins (Gallardo et al., 2002). Interestingly, 22 LEA genes were found to be highly expressed at the transcript level in vegetative tissues, in particular under dehydration or exposure to low temperature (Hundertmark and Hincha, 2008). Indeed, several LEA proteins are...
known as cold responsive proteins, such as COLD RESPONSEV15A (COR15A) in Arabidopsis (Steponkus et al., 1998).

No enzymatic activity has been reported so far for LEA proteins, which are more likely expected to play a role in stress tolerance as protectants of biomolecules and membranes. As intrinsically disordered proteins, they are proposed to have a broad impact in the abiotic stress response in plants (Sun et al., 2013). Individual LEA proteins were shown in vitro to protect enzymes from desiccation and/or freezing induced aggregation (Goyal et al., 2005b; Grelet et al., 2005; Reyes et al., 2005; Pouchkina-Startceva et al., 2007; Nakayama et al., 2008; Sharon et al., 2009; Boucher et al., 2010). Such protection of poly pep tides might result from chaperone-type effects, implying direct interaction of LEA proteins with clients to prevent aggregation (Kovacs et al., 2008), or from space-filling by LEA proteins, which reduces the collision rates between aggregating proteins, the so-called molecular shield function (Chakrabortee et al., 2012). LEA proteins from the dehydrin family have been shown to bind membranes through their canonical K-segment, a Lys-rich peptide (Koag et al., 2003, 2009). The phosphorylation-dependent binding of the K-segment to phospholipid vesicles was later shown to lower membrane lipid phase transition, in agreement with a role of dehydrin in cold tolerance (Eriksson et al., 2011). A pea (Pisum sativum) seed mitochondrial LEA protein of the LEA_4 family was shown to protect membranes from desiccation (Tolletter et al., 2007, 2010). This intrinsically disordered protein, located in the matrix space, folds during dehydration into a class A amphipathic α-helix form that inserts laterally in the membrane to afford protection in the dry state. LEA proteins have also been shown to sequester calcium (Alsheikh et al., 2005), metal ions (Svensson et al., 2000; Kruger et al., 2002), and reactive oxygen species (Hara et al., 2004), thus potentially contributing to dehydration and cold stress tolerance. They have also been shown to contribute to the glassy state, which is formed by nonreducing sugars such as sucrose or trehalose upon desiccation in the cytoplasm of most of anhydrobiotes (Wolkers et al., 2001; Shih et al., 2004; Shimizu et al., 2010).

Ectopic expression of individual LEA proteins in various organisms has also provided support for a function in stress protection, although the effects are often moderate (for review, see Tunnatilfe and Wise, 2007). For instance, a brown mustard (Brassica juncea) dehydrin overexpressed in yeast (Saccharomyces cerevisiae) or tobacco (Nicotiana tabacum) increased salt and cold tolerance in both organisms (Xu et al., 2008). Another wheat (Triticum aestivum) LEA protein was shown to increase cold tolerance when overexpressed in Arabidopsis (NDong et al., 2002). Recently, a dramatic effect was reported for a brine shrimp (Artemia franciscana) mitochondrial LEA protein expressed in cultured human hepatoma cells (Li et al., 2012b). Indeed, membrane integrity of transgenic cells was preserved after a rapid and uniform cell desiccation, and with the combination of intracellular loading of trehalose before drying, cells maintained almost normal proliferation rates after rehydration.

The diversity of LEA proteins in eukaryotes raises the question of their subcellular localization, and although LEA proteins are predicted to populate most cellular compartments, relatively few have been experimentally localized (Tunnaciffe and Wise, 2007). LEA proteins have been found mostly in the cytosol (Mundy and Chua, 1988; Franz et al., 1989; Roberts et al., 1993; Goyal et al., 2005a) and in chloroplasts (Iturriaga et al., 1992; Lin and Thomashow, 1992; NDong et al., 2002). LEA proteins have also been identified in mitochondria from pea (Bardel et al., 2002; Grelet et al., 2005) and from the brine shrimp (Menze et al., 2009). LEA proteins have also been localized in Arabidopsis peroxisomes (Cutler et al., 2000) and in the endoplasmic reticulum (ER) (Ukai et al., 2001). In the bdelloid rotifer Adineta viccis, two LEA proteins were localized in the ER and shown to be distributed to vesicles and partially secreted (Tripathi et al., 2012). Several LEA proteins have been shown to accumulate in the nucleus (Borrell et al., 2002) or to be translocated from the cytosol to the nucleus upon phosphorylation (Jensen et al., 1998; Riera et al., 2004).

For the model plant Arabidopsis, with 51 LEA genes, information about the subcellular localization of the corresponding proteins can be mined in the SUBA (Arabidopsis Subcellular Database) (Heazlewood et al., 2007) and MASCP Gator (Joshi et al., 2011) databases. Experimental information derived from subcellular proteomics or fluorescent protein translational fusions is only available for 16 of the Arabidopsis LEA proteins, and among those, four proteins display several different localizations.

Considering LEA protein diversity, and their major, yet still enigmatic, functions with respect to desiccation tolerance, and possibly other stress responses, establishing the cellular location of each protein is critical to our understanding of their precise biological function. Here, we examined the subcellular localization of the 51 Arabidopsis LEA proteins using a combination of bioinformatic analysis and experimental localization of translational protein fusions. The data revealed their wide distribution and the potential role of the largest family in the protection of the various cellular membranes.

RESULTS

This work aimed to resolve the subcellular localization of the LEA proteins encoded by the 51 LEA genes identified by Hundertmark and Hincha (2008). For the sake of clarity, we will use their nomenclature, which numbers genes successively on the genome. Table 1 shows the correspondance with the AGI code of the protein, alternative names of the proteins, and their classification using PFAM and LEAPdb. For LEA genes with two gene models in The Arabidopsis Information Resource (http://www.arabidopsis.org/), we selected the protein from the gene model with the highest number of ESTs.

Bioinformatics Subcellular Predictions

A bioinformatics approach was developed in order to predict the localization of the 51 proteins. Among the several dozen subcellular prediction programs available on the Internet (see http://www.psort.org/), we selected a panel of eight that were able to provide localization probabilities for one or several compartments, including plastid, and thus trained for plant proteins. Two other programs (BaCello and WolF PSORT), which offer predictions for multiple compartments, but without scores, were included because of their reliability (Casadio et al., 2008). All the results obtained with the first eight programs giving prediction scores have
<table>
<thead>
<tr>
<th>Arabidopsis LEA Protein Nomenclature and Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene Model</strong></td>
</tr>
<tr>
<td><strong>AGI Code</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>21</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>28</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>33</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>36</td>
</tr>
<tr>
<td>37</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>44</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>46</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>51</td>
</tr>
</tbody>
</table>

aData obtained from SUBA database, corresponding to either GFP or tandem mass spectrometry experiments. Numbers indicate the amount of studies showing same localization. C, cytosol; N, nucleus; Pm, plasma membrane; P, plastid; Perox, peroxisome; V, vacuole; G, Golgi apparatus.
been gathered in Figure 1 as a heat map and organized by compartments. The results suggest a rather high representation of LEA proteins in the cytosolic and nuclear compartments, but reveal a lot of heterogeneity in the predictions. If we apply an arbitrary cutoff by considering as valid only predictions for which more than half of the algorithms give scores equal to or over 50% (i.e., three out of the four programs used for these compartments), the heat map appears less scattered (Supplemental Figure 1). The predictions by BaCelLo and WoLF PSORT, which do not provide scores, match poorly with the other predictors (Supplemental Figure 1). In fact, the programs were unanimous in giving high scores for a single localization for only two out of 51 proteins, LEA21 and LEA22, which are expected to populate the endomembrane system (ES). In summary, this bioinformatics analysis, for which up to 10 algorithms were applied to the 51 polypeptides, highlights the difficulty in predicting LEA protein localization with high confidence.

**Expression of Fluorescent LEA Protein Fusions in Protoplasts, Seedlings, and Plants**

In order to acquire experimental data about the subcellular distribution of the 51 LEA proteins, the corresponding coding sequences were genetically fused with green or red fluorescent proteins (GFP and RFP) in vectors designed for transient transgene expression in Arabidopsis leaf protoplasts, which were then observed by laser scanning confocal microscopy. GFP constructs were designed as both N- or C-terminal fusions to each LEA protein. Subcellular compartments markers were used in coexpression experiments to validate the different locations when appropriate and to confirm some peculiar locations. In addition, six constructs were also expressed in seedlings, using a transient Agrobacterium tumefaciens-mediated transformation of Arabidopsis, to examine their location in a tissue context, and transgenic lines were generated to confirm the localization of 11 proteins.

**Cytosol and Nucleus**

For the majority of LEA proteins (36), both the LEA-GFP and the corresponding GFP-LEA protein fusions were routinely observed to give a diffuse fluorescent pattern typical of a cytosolic location, clearly excluding chloroplasts and vacuoles. Since it is difficult to determine with Arabidopsis protoplast whether cytosolic protein fusions also enter the nucleus, like GFP alone, all of these LEA-GFP constructs were coexpressed with a nuclear localization signal (NLS)-RFP, which is exclusively targeted to the nucleus. Figure 2 shows merged images of the colocalization of NLS-RFP and LEA-GFP fluorescence for a series of four typical LEA proteins exhibiting

![Figure 1. Heat Map of Subcellular Localization Predictions for the 51 Arabidopsis LEA Proteins.](image)

The heat map illustrates the probability, given by the different programs indicated on the right, of LEA protein targeting to various cellular compartments. LEA proteins are numbered from 1 to 51, and the color coding represents the targeting probability from 0 (yellow) to 100% (red).
an exclusive cytosolic localization. The images in Figure 2 show also another series of four LEA proteins displaying a dual localization in the cytosol and nucleus. Among these 36 LEA protein fusions, seven were found exclusively cytosolic, while the others were confirmed with a dual cytosolic-nuclear localization (Table 2). Images captured in separate channels and using the inverted (GFP-LEA) fusion for all these constructs are shown in Supplemental Figure 2. Dual cytosolic-nuclear localization could result from passive diffusion of the fusion proteins into the nucleus or from the presence of degraded LEA-GFP having lost the NLS but still remaining fluorescent. Plotting the repartition of GFP constructs between cytosolic or cytosolic-nuclear localizations according to their molecular mass clearly reveals that dually localized proteins have smaller molecular masses than the cytosolic constructs, which are excluded from the nucleus (Supplemental Figure 3). These data fit well with the classical estimation of the nuclear exclusion limit around 60 kD (Dingwall and Laskey, 1986), and passive diffusion in the nucleus of highly expressed cytosolic LEA-GFP would explain their dual cytosolic-nuclear localization. The integrity of the constructs was verified by protein gel blotting of protoplast protein extracts using an antibody directed against GFP (Supplemental Figure 4). Although some degradation products can be detected, a major band with a SDS-PAGE apparent molecular mass close to, or a few kilodaltons higher than, the theoretical size of the protein fusions is always observed (Supplemental Table 1). Such increased apparent molecular mass on SDS-PAGE gels could be attributed to the disordered character of LEA poly-peptides, which leads to reduced binding of SDS (Tompa, 2002).

In particular, two cytosolic-nuclear constructs (LEA31-GFP and LEA32-GFP) with respective theoretical masses of 55.370 and 55.405 kD (Supplemental Figure 3) approaching the nuclear exclusion limit displayed strong signals with a very low level of degradation products (Supplemental Figure 4). This indicates that these intact constructs still diffuse passively in the nucleus. Conversely, LEA10-GFP, which has the lower theoretical molecular mass (49.412 kD) of the cytosolic protein group, i.e., below the nuclear exclusion limit, does not diffuse into the nucleus. However, it displays a much higher apparent molecular mass of 63 kD (Supplemental Figure 4 and Supplemental Table 1), and its structural features might prevent passive diffusion through nuclear pores. Taken together, these experiments indicate that 36 LEA-GFP constructs likely accumulate in the cytosol, while 29 of them are able to diffuse passively in the nucleus.

**Mitochondria and Plastid**

Nine LEA-GFP fusions were found to be targeted to mitochondria, plastid, or to both organelles. LEA37, 38, and 41 fusions displayed a typical mitochondrial matrix pattern illustrated by the perfect colocalization between the GFP signal and the mitochondrial mCherry marker (Figure 3). Interestingly, two fusion proteins (LEA42 and LEA48) were found to be dually targeted to mitochondria and plastids (Figure 3). As a control, GFP-LEA fusions, in which GFP should mask the mitochondrial targeting peptide, were observed in the cytosol as a diffuse signal (Supplemental Figure 5). Figure 4 illustrates the localization of four LEA-GFP proteins (LEA11, 12, 23, and 24) that were found exclusively in plastids, where the autofluorescence of the chlorophyll was used as a marker. Like in the case of the mitochondrial LEA proteins, masking the N terminus targeting peptides (GFP-LEA fusions) resulted in diffuse cytosolic localizations (Supplemental Figure 5). For the five LEA protein fusion targeted to mitochondria or dual targeted to plastid and mitochondria, we established transgenic lines. Confocal examination confirmed the dual targeting of LEA42 and LEA48 and the exclusive mitochondrial localization of LEA 37, 38, and 41 (Supplemental Figure 6).

For the fusion proteins localized in plastid or dual targeted, the verification of fusion protein expression by protein gel blot analysis of protoplasts (Supplemental Figure 4) revealed major bands with

---

**Table 2. Cytosolic and Nuclear LEA Proteins**

<table>
<thead>
<tr>
<th>Localization</th>
<th>LEA Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosol</td>
<td>4, 5, 10, 19, 25, 26, 36</td>
</tr>
<tr>
<td>Cytosol and</td>
<td>1, 2, 3, 6, 7, 8, 14, 15, 16, 17, 18, 20, 27, 28, 29, 31</td>
</tr>
<tr>
<td>Nucleus</td>
<td>32, 33, 34, 35, 39, 40, 44, 45, 46, 47, 49, 50, 51</td>
</tr>
</tbody>
</table>

---

**Figure 2.** Transient Expression of Cytosolic LEA Protein in Arabidopsis Protoplasts.

Representative examples of LEA-GFP fusion proteins located either in the cytosol only or both in the cytosol and the nucleus were coexpressed with a RFP nuclear marker in Arabidopsis mesophyll protoplasts. Numbers refer to the corresponding LEA proteins. Green, GFP; red, RFP; blue, chlorophyll. The other LEA proteins with similar locations are illustrated in Supplemental Figure 2. Bars = 10 μm.
SDS-PAGE apparent molecular masses lower than theoretical sizes of the protein fusions (Supplemental Table 1). These few kilodaltons of differences would be in agreement with the cleavage of targeting peptides during translocation. In the case of the three proteins (LEA37, 38, and 41) exclusively targeted to mitochondria, the differences between theoretical masses of the fusion proteins and their apparent size in protein gel blots were not convincing with respect to the cleavage of targeting peptides. According to protein gel blots, LEA37 fusion was found slightly smaller (0.5 kD), LEA41 almost equal, and LEA38 slightly bigger (1 kD) than their respective theoretical size (Supplemental Figure 4 and Supplemental Table 1). However, because of the small size of these LEA proteins (10 to 14 kD), one cannot rule out the possibility that these proteins would be imported with a very short targeting peptide or even without any cleavable sequence.

**ER, Vacuole, and Secretion**

Three LEA-RFP fusion proteins were found to populate the ER in transformed protoplasts, as illustrated in Figure 5. LEA13 and LEA30 signals appeared to clearly colocalize with a GFP-ER marker. When LEA30-RFP was transiently expressed in a seedling system, in order to examine subcellular distribution in intact tissues, the protein was also found to colocalize with the ER marker (Figure 5). In the case of LEA13-RFP, bright red dots, distinct but close to the ER, were observed in protoplasts, suggesting an additional localization in Golgi. However, we could not observe colocalization with a Golgi marker (Supplemental Figure 7). When expressed in seedlings, LEA13-RFP was surprisingly found to accumulate in the vacuole (Figure 6), which could be the final destination of the protein in the secretory pathway. In order to gain more information about the targeting of these proteins, we established transgenic lines, which confirmed the vacuolar localization for LEA13-RFP and showed that LEA30-RFP was localized both in the ER, as in protoplasts and seedlings, but also in the vacuole (Supplemental Figure 8). In protoplasts, LEA43-RFP was found to accumulate specifically in several ovoid regions of the ER (Figure 5). In transiently transformed seedlings, LEA43-RFP accumulated in large and rather diffuse areas of the ER (Figure 6), possibly corresponding to ER sheets, but not in areas known as fusiform bodies (Hawes et al., 2001). In the transgenic lines overexpressing LEA43-RFP, the protein was found to reside in the ER, like in protoplasts and seedlings (Supplemental Figure 8).

Constructions with GFP fused upstream of LEA proteins, in order to mask a putative N terminus signal peptide, were transformed into protoplasts. As expected, all three GFP-LEA fusions were found to reside in the cytosol with a diffuse pattern (Supplemental Figure 9).

To check the integrity of the fusion proteins, protoplasts were transformed with these LEA-GFP constructs and subjected to protein gel blotting. The three proteins fusions were clearly detected with apparent molecular masses close to their theoretical mass (Supplemental Figure 4 and Supplemental Table 1). However, considering the size of the fusions proteins (61 to 78
kDj, it is difficult to reach any conclusion regarding the cleavage or lack of a putative signal peptide.

Two LEA fusions (LEA21-RFP and LEA22-RFP) were identified as vacuolar proteins in protoplasts expressing the ER-GFP marker (Figure 7). LEA21-RFP was essentially found in the vacuole (Figure 7), with additional bright spots that could possibly correspond to Golgi (Supplemental Figure 10). LEA22-RFP was observed to accumulate both in the vacuole and in the ER (Figure 7). To confirm the targeting of these two proteins to the vacuole via the secretory pathway, a treatment with Brefeldin A (BFA) was performed to block ER-Golgi transport (Nebenführ et al., 2002) (Figure 7). In BFA-treated protoplasts, LEA21-RFP clearly did not reach the vacuole and was instead found to accumulate in the ER and in large vesicles (Figure 7). These could possibly correspond to ER-Golgi transport vesicles stacking next to the ER exit site and unable to reach the BFA impaired cis-Golgi (Nebenführ et al., 2002). Vacuole accumulation of LEA22-RFP was also prevented by BFA, and the protein instead remained mostly associated with ER (Figure 7). This is in agreement with their ER and vacuolar localization, with BFA blocking the secretion to the vacuole without hampering ER accumulation. To check the integrity of the fusion proteins, protoplasts were transformed with the respective LEA-GFP constructs. In agreement with the fact that GFP targeted to vacuole is readily degraded (Tamura et al., 2003), the corresponding LEA-GFP constructs were not detected in the vacuole of transformed protoplasts, but only as bright spots for LEA21-GFP and ER-type network for LEA22-GFP, although fluorescence was very weak for that construct (Supplemental Figure 11). In protein gel blots, main bands with an SDS-PAGE apparent molecular mass in the range of the theoretical size of the protein fusions could nevertheless be detected for the two LEA proteins (Supplemental Figure 4 and Supplemental Table 1).

The two LEA-RFP fusion proteins were also transiently expressed in seedlings. Both LEA21 and LEA22 fusions were found to accumulate in diffuse areas in the intercellular space surrounding transformed cells and hence are likely secreted (Figure 6). While LEA21-RFP appears exclusively secreted in these conditions, LEA22-RFP was also observed within the ER and the vacuole. GFP-LEA controls were found localized in the cytosol with a diffuse fluorescence, which suggests that putative N terminus signal peptides would be masked (Supplemental Figure 9). In leaf cells from a transgenic line expressing LEA21-RFP, the protein was found to be clearly secreted in the intercellular space, while fluorescence was also observed in vacuole (Supplemental Figure 8). In the case of transgenic lines expressing LEA22-RFP, the fusion protein was found to accumulate in the vacuole, but not in the ER (Supplemental Figure 8).

**Pexophagosome**

Interestingly, when expressed in protoplasts, LEA9-RFP displayed an unusual pattern that could not be attributed to any classical compartment (cytosol, nucleus, ER, Golgi, vacuole, peroxisome, mitochondria, and plastid). The protein accumulates in peculiar spherical and apparently hollow structures (Figure 8). When LEA9-RFP was expressed in cells expressing an ER-GFP marker, these structures appeared to neighbor or partially overlap the ER, suggesting a close association (Figure 8). The coexpression of LEA9-RFP with a peroxisomal cyan fluorescent protein (CFP) marker revealed that the LEA9 fusion populated structures that engulfed most peroxisomes (Figure 8). These unusual structures could therefore correspond to pexophagosomes, which are organelles involved in the recycling of...
peroxisomes (van Zutphen et al., 2008). Analysis of the integrity of the fusion protein (Supplemental Figure 4) in protoplasts indicates that the apparent molecular mass corresponds to the theoretical size of the protein fusions (Supplemental Table 1). In order to confirm this location, LEA9-RFP was expressed in seedlings where it displayed a pattern typical of pexophagosomes (Figure 8). Transgenic lines expressing LEA9-RFP and a peroxisome marker were established, but few cells were found to accumulate LEA9-RFP. However, the protein always displayed a similar pattern, with association with peroxisomes (Supplemental Figure 8). Finally, a construct with an opposite orientation (GFP-LEA9) expressed in protoplasts displayed similar images to those of its LEA-RFP counterpart (Supplemental Figure 9), suggesting that targeting information was not masked by GFP and, hence, not located in the N or C terminus of the protein.

**Targeting Peptide Cleavage Sites for Mitochondrial and Plastid LEA Proteins**

Proteins targeted to mitochondria and plastids generally carry an N-terminal targeting peptide, which is cleaved during the import process (Teixeira and Glaser, 2013). As a result, the mature proteins released inside the organelle are usually shorter than their cytosolically synthesized precursors. To facilitate the proper structural and functional characterization of organellar proteins, it is critical to determine the targeting peptide cleavage sites. We previously described a method to experimentally determine the targeting peptide cleavage site of C-terminal GFP fusion proteins imported into mitochondria or chloroplasts during transient expression in protoplasts (Candat et al., 2013). In this approach, protoplasts expressing LEA-GFP proteins are lysed and fusion proteins are immune-captured with GFP antibodies bound to magnetic microbeads. Purified protein-GFP fusions are then separated by electrophoresis and subjected to Edman microsequencing to determine their N terminus. Table 3 displays the targeting peptide cleavage sites we were able to determine for LEA proteins identified in plastids or mitochondria. The cleavage sites for LEA23-GFP and LEA24-GFP, which are localized in plastids, were determined previously (Candat et al., 2013). In the case of LEA11 and LEA12, which are two paralogous proteins targeted to plastids (Figure 4), the cleavage site could be only be determined for LEA11-GFP (Table 3). The level of expression of the LEA12-GFP construct in protoplasts was too low for successful N terminus microsequencing. However, alignment of LEA11 and LEA12 polypeptide sequences, shown in Supplemental Figure 12, indicates a high conservation of the cleavage site regions SWV(S/P)(AVKG(A/D)GNS (arrow indicates cleavage site). It is therefore highly probable that cleavage of the LEA12 precursor occurs at the proposed site. In the case of LEA42 and LEA48, which are dually targeted to plastids and mitochondria (Figure 3), a unique cleavage site was identified (Table 3). This suggests that these proteins should carry an ambiguous transit peptide targeting the precursors to both mitochondria and plastids. We were unfortunately unable, so far, to determine a targeting peptide cleavage site for the three LEA proteins (LEA37, 38, and 41) identified in mitochondria (Figure 3). As in the case of LEA12, their relatively low level of expression in protoplasts hampers the analysis.

**Figure 7.** Effect of BFA on ER- or Vacuole-Targeted LEA Proteins Transiently Expressed in Arabidopsis Protoplasts.

LEA-RFP fusion proteins localized in the ER or the vacuole were coexpressed with an ER marker (ER-GFP) in Arabidopsis mesophyll protoplasts. When indicated, transient expression was performed in the presence of BFA (+BFA). Numbers refer to the corresponding LEA proteins. Green, GFP; red, RFP. Bars = 10 µm.

**Figure 8.** Transient Expression of a Pexophagosome-Targeted LEA Protein in Arabidopsis Protoplasts or Seedlings.

LEA9-RFP fusion protein localized in pexophagosomes was coexpressed with an ER marker (ER-GFP) or a peroxisome marker (Perox-CFP) in Arabidopsis mesophyll protoplasts or in Arabidopsis seedling cotyledons using Agrobacterium. Green, GFP or CFP; red, RFP; blue, chlorophyll. Bars = 10 µm for protoplasts and 20 µm for seedlings.
DISCUSSION

LEA proteins are mostly intrinsically disordered proteins expected to play major roles in abiotic stress tolerance in anhydrobiotes. In plants, several dozen genes encode LEA proteins that can be clustered into several families according to their sequence properties. The aim of this research was to provide a subcellular map of the distribution of LEA proteins in the model plant Arabidopsis, for which 51 LEA genes were inventoried (Hundertmark and Hincha, 2008). Because of the high number of gene fusion constructs to investigate (and both N- and C-terminal fusions), a leaf mesophyll protoplast transient assay was selected (Yoo et al., 2007). Although observations performed with mesophyll protoplasts are not representative of all cell types and tissues, they provide a robust system to analyze the intrinsic routing information carried by polypeptides, especially in a comparative manner as in this study. In addition, for most LEA proteins not displaying a cytosolic localization in protoplasts, complementary approaches of Agrobacterium-mediated transient expression in seedlings or stable transgenic lines were used.

Comparison between Subcellular Localization Predictions and Experimental Data

The subcellular localization of more than 50 related proteins, as experimentally determined in this work, represents an interesting data set to examine the performance of different targeting prediction algorithms. A schematic overview of the relative performance of eight different software packages is presented in Supplemental Figure 13. The bar graph illustrates the proportion of correct and erroneous predictions by the different programs for ES, mitochondria, plastids, and cytosolic compartments. For the latter, we considered that all proteins dually localized in the cytosol and nucleus were actually cytosolic and only diffused passively into the nucleus. Since we could not demonstrate exclusive nuclear localization for any construct, all the many nucleus targeting predictions were false and were not included in the graphical analysis. Clearly, the predictions for ES were more accurate than for other compartments, with only three erroneous predictions out of a total of 34, with five programs yielding a high success rate (at least 5% for YLoc and much lower for the other programs). In the case of plastid and mitochondrial compartments, while the success rate was generally good, the proportion of erroneous predictions was high (Supplemental Figure 13). This certainly reflects the insufficient knowledge about the targeting peptide information for these organelles, as well as the similarity in their import machineries and the increasing number of proteins shown to be dual targeted (Carrie et al., 2009). Overall, PProwler was the most efficient predictor, followed by YLoc and TargetP. However, this conclusion only applies to our data set of proteins, which are highly hydrophilic, largely disordered, and composed of a majority of cytosolic proteins. Nevertheless, this study confirms that assigning protein localization is clearly strengthened by the use of multiple programs, as was demonstrated by analysis of the plant mitochondrial proteome (Heazlewood et al., 2004).

Subcellular Distribution and Classification of LEA Proteins

The experimental data enables an examination of the subcellular distribution of the Arabidopsis LEA proteome with respect to the structural classification of the proteins into several families, which likely have different evolutionary histories. Figure 9 shows a cladogram of the 51 LEA proteins incorporating their PFAM classification and their subcellular location. Among the nine families, six (LEA_1, LEA_2, LEA_5, LEA_6, dehydrin, and SMP) display a cytosolic (or cytosolic and nuclear) localization for all their members (up to 10 in the case of the Dehydrin family). The two members of the A M family share localization within the secretory pathway, while three out of four members of the LEA_3 family are mitochondrial, the last being cytosolic (Figure 9). In contrast to the rather homogenous location of members of those families, the 18 members of the LEA_4 family, which is the largest family, displayed multiple localizations (cytosol, plastid, mitochondrion, ER, and pexophagosome). This overall analysis clearly highlights a link between protein family and location, with the members of eight families generally displaying a single common location while the ninth family is distinguished by members having targeted to diverse locations.

Cytosolic and Nuclear LEA Proteins

The vast majority (36, distributed in six families) of Arabidopsis LEA protein appeared cytosolic and able to diffuse passively in the nucleus. It is noteworthy that we could not observe any LEA fusion proteins specifically targeted to the nucleus, although LEA31 was previously assigned a nuclear localization using GFP constructs (Borrell et al., 2002). However, from the examination of the published images of Arabidopsis RESPONSIVE TO ABSCISIC ACID28 (AtRab28 and LEA31) stably expressed as a GFP fusion in Arabidopsis (Borrell et al., 2002), the fusion protein appears in fact to be distributed in both the cytosol and nucleus.

<table>
<thead>
<tr>
<th>LEA Protein</th>
<th>Localization</th>
<th>Experimental Determination of the Mature Protein N-Terminal End (Edman)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Plastid</td>
<td>50-51 AVKGAG</td>
</tr>
<tr>
<td>12</td>
<td>Plastid</td>
<td>ND</td>
</tr>
<tr>
<td>23</td>
<td>Plastid</td>
<td>51-52 VKSDGN</td>
</tr>
<tr>
<td>24</td>
<td>Plastid</td>
<td>49-50 AAKGDG</td>
</tr>
<tr>
<td>42</td>
<td>Mitochondria and plastid</td>
<td>40-41 TSVSQN</td>
</tr>
<tr>
<td>48</td>
<td>Mitochondria and plastid</td>
<td>41-42 SSVNHS</td>
</tr>
</tbody>
</table>

ND, not determined.

* Determined previously (Candat et al., 2013).
Three other proteins (LEA4, LEA5, and LEA8) were previously described as cytosolic using GFP fusions (Cutler et al., 2000; Abu-Abied et al., 2006), which we could confirm. Two others (LEA10 and LEA46) are tentatively identified as peroxisomal according to the SUBA database, based on the results of random GFP-cDNA expression in Arabidopsis (Cutler et al., 2000). However, such results should be treated with caution due to the high proportion of non-native coding sequences (i.e., out-of-frame cDNAs) in this large-scale analysis, which resulted in peroxisomal-like localization. Two proteins (LEA7 and LEA29) that appeared cytosolic in our experiments were proposed to reside mainly in the ER, based on GFP fusions expressed in protoplasts (Zhao et al., 2011). However, using colocalization with the ER-GFP marker (Supplemental Figure 14), we could not find any evidence for an ER location for LEA7 or LEA29 GFP fusions.

The increasing amount of subcellular proteomic data that are available in public databases (e.g., MASCOT Gator and SUBA) allows a comparison of our results with those obtained using other approaches. Among the LEA set of cytosolic proteins, six (LEA1, 4, 5, 10, 26, and 27) were identified in the cytosolic proteome of Arabidopsis cell suspensions (Ito et al., 2011). Five of them (LEA4, 5, 10, 26, and 27), as well as two others (LEA8 and LEA16), were identified in the plasma membrane proteome of cell cultures or leaves (Alexandersson et al., 2004; Benschop et al., 2007; Marmagne et al., 2007; Mitra et al., 2009; Keinath et al., 2010; Elmore et al., 2012; Li et al., 2012a; Nikolovski et al., 2012). Five proteins (LEA4, 5, 8, 10, and 33) belonging to the dehydrin family could interact with the plasma membrane since LEA5, LEA10, and LEA33 have been shown to bind anionic phospholipid vesicles (Kovacs et al., 2008; Eriksson et al., 2011). Moreover, dehydrins from maize (Zea mays) and the salt-tolerant Thellungiella salsuginea have also been shown to bind phospholipid vesicles (Koag et al., 2003, 2009; Rahman et al., 2010, 2013). Interestingly, four dehydrins (LEA4, 5, 8, and 10) with a clear cytosolic localization in our experiments were also identified in chloroplast proteome (Kleffmann et al., 2004; Zybalov et al., 2008; Ferro et al., 2010; Kong et al., 2011) or in the tonoplast in the case of LEA10 (Whiteman et al., 2008). This suggests interaction of these cytosolic dehydrins with other cellular membranes. LEA26, another cytosolic protein of the LEA_2 family identified in cytosolic and plasma membrane proteomes (see above), was also identified in the Golgi proteome (Parsons et al., 2012). This protein, for which little information apart from its downregulation following cadmium exposure of Arabidopsis cell culture (Sarry et al., 2006) is available, could possibly also interact with membrane. Finally, LEA29, whose localization is debated above, was identified in the nuclear proteome of Arabidopsis leaves (Bae et al., 2003), which agrees well with its capacity to diffuse in the nucleus.

Overall, comparison between our data and that arising from subcellular proteomics indicates discrepancies for several proteins that are in fact easily explicable: for example, localization in the nucleus as a result of diffusion rather than targeting or association with various compartments as a result of simple contamination or membrane binding. Also, the lack of detection of many LEA proteins in subcellular proteomics is likely be due to the absence of transcription or low expression of the corresponding LEA genes.

The Plant Cell

Figure 9. Cladogram Illustrating the Repartition of Arabidopsis LEA Protein Subcellular Distribution in Each Family.

The rectangular cladogram was built with Dendroscope software (see Methods) using a multiple alignment of the 51 LEA proteins, which are indicated by their number. The subcellular localization is indicated by letters on the right. C, chloroplast; E, ER; M, mitochondria; P, pexophagosome; V, vacuole. All proteins without indication are cytosolic. The brackets on the right indicate the different LEA protein families. DHN, dehydrin.

AtM LEA Proteins Are Vacuolar or Secreted

Two related LEA proteins (LEA21 and LEA22) could be specific to Brassicaceae and were previously classified in the AtM family (Hundertmark and Hincha, 2008). In our study, the two AtM LEA fusions LEA21- and LEA22-RFP appeared clearly in the ES but
displayed various localizations (ER, vacuole, and secretion) depending on the expression system, with, for instance, excretion being undetectable with protoplasts. It is worth noting that these two proteins were only detected in seeds (Supplemental Figure 15) in the course of a comprehensive organ-specific proteomics analysis in Arabidopsis (Baerenfaller et al., 2011). This suggests a role for these LEA proteins, with respect to their ER and vacuolar localization, in the biogenesis and degradation of protein storage vacuoles (Otegui et al., 2006). If the proteins are actually secreted in seeds, they would be expected to have a role as extracellular chaperones in dehydration tolerance within the intercellular spaces, as was suggested for two LEA proteins from an anhydrobiotic bdelloid rotifer (Tripathi et al., 2012).

Mitochondrial and Cytosolic LEA Proteins in the LEA_3 Family

Three of the four LEA proteins (LEA37, LEA38, and LEA41) comprising the LEA_3 family were exclusively targeted to mitochondria, which was confirmed also in transgenic lines, while the remaining protein (LEA2) was ascribed a cytosolic localization. Interestingly, LEA2 and LEA38 are two paralogous proteins (Hundertmark and Hincha, 2008), which are targeted to different locations. Although the two sequences are very similar, the location differences could be explained by a deletion of six amino acids (36-ValSerSerGlyGlyArg-43) in LEA2, equating to a region prior to the putative cleavage site of LEA38 (Supplemental Figure 16). LEA38 was recently identified in mitochondria, which is consistent with our observations, and proposed to play major roles in plant development as well as abiotic and biotic stress responses (Salleh et al., 2012). Since the protein was previously shown to improve oxidative stress tolerance when expressed in yeast (Mowla et al., 2006), an interaction of LEA38 with enzymes involved in mitochondrial reactive oxygen species metabolism and signaling was suggested to mediate the pleiotropic effects observed in overexpressor or antisense Arabidopsis lines (Salleh et al., 2012). The two other mitochondrial proteins (LEA37 and 41), as well as the cytosolic LEA2, share similar sequence properties, which would suggest analogous molecular functions (albeit in a different compartment for LEA2).

In the organ-specific proteomics mapping, while LEA2 and LEA41 were detected exclusively in flowers, and LEA38 in flower and roots, LEA37 was not detected in any compartment (Supplemental Figure 15). Although the levels of expression of these genes at the transcript level were very low in the publicly available data (EFP browser: http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi), this does not preclude an increase under particular stress conditions.

Multilocalization within the LEA_4 Family

A striking feature of the LEA_4 family in Arabidopsis is the diversity of subcellular locations observed for the 18 corresponding LEA-GFP or LEA-RFP proteins. While eight proteins appeared cytosolic, five were identified in chloroplasts, two in mitochondria and chloroplasts, three in the ER, and one in structures identified as pexophagosomes. Among the four protein fusions targeted to chloroplast in this study, LEA23 was previously identified in chloroplast proteomic analyses (Froehlich et al., 2003; Pettier et al., 2006; Zybailov et al., 2008; Ferro et al., 2010) and LEA24 was shown to be plastidic by transient expression in protoplasts (Candat et al., 2013). However, LEA12, which shows a clear chloroplastic location as a GFP fusion, was previously ascribed a plasma membrane location (Mitra et al., 2009), which could possibly result from contamination by the cytosolic precursor. Organ-specific proteomics indicates that LEA11 and LEA12 are only detected in flowers, while LEA23 and LEA24 are abundant in flowers, leaves, and siliques, but are not detected in seeds (Supplemental Figure 15). These data do not support a general role for plastid LEA proteins in the desiccation tolerance of seeds. However, these proteins could be expressed at higher abundances under stress conditions. Accordingly, LEA23 and LEA24 are known as cold-regulated proteins (COR15B and COR15A) involved in the stabilization of chloroplast membranes or enzymes during freezing (Steponkus et al., 1998; Nakayama et al., 2008; Thalhammer et al., 2010). Two paralogous LEA proteins (LEA42 and LEA48) appeared dual targeted to mitochondria and chloroplasts, both in protoplasts and transgenic lines. They are likely orthologous to LEAM, previously identified in mitochondria from pea seeds (Bardel et al., 2002; Grelet et al., 2005) and from brine shrimp encysted embryos (Menze et al., 2009). Since LEAM-GFP was not found to be targeted in the chloroplast in transformed pea protoplasts (Grelet et al., 2005), dual targeting of these LEA proteins to mitochondria and plastids might not be a general property in plants. It is expected that LEA42 and LEA48 play roles similar to that of LEAM with respect to membrane protection during desiccation (Tollet er et al., 2007), which is in agreement with their seed-specific expression illustrated by their organ-specific proteomics (Supplemental Figure 15) or transcript abundance (EFP browser: http://bar.utoronto.ca/ efp/cgi-bin/efpWeb.cgi).

Three proteins (LEA13, LEA30, and LEA43) were found to engage in the ES pathway. LEA13 was previously detected in a plasma membrane proteomics analysis (Mitra et al., 2009), which may be consistent with transit through the ES since vacuole targeting and secretion through the plasma membrane are correlated. For these three proteins, which reside or transit through the ER, putative retention signals resembling the canonical KDEL sequence were detected in the C terminus (LEA13: YAEL; LEA30: DAEL; LEA43: SAEI). However, ER retention signals can be very specific since a single substitution can abolish their function; moreover, they can be passenger dependent (Denecke et al., 1992; Vitale and Denecke, 1999). Therefore, a more detailed analysis of these putative signals, in particular for LEA13, including the design of constructs with a central fluorescent protein, would be required to ascertain their function. According to the organ-specific proteomics analysis, while these ER proteins can be detected in flowers, leaves, or pollen, LEA30 and LEA43 are especially abundant in seeds, suggesting a role in desiccation tolerance (Supplemental Figure 15).

LEA9, which displays a seed-specific expression pattern (Supplemental Figure 15), was the sole LEA protein that was tentatively identified in pexophagosomes. This raises the question of the existence of pexophagosomes in seeds. Since glyoxysomes and peroxisomes are required for the conversion of lipid reserves during seed germination and the transition to photoautotrophy in Arabidopsis (Graham, 2008), pexophagosomes could have an
important role with respect to the biogenesis and recycling of these organelles. LEA9 could therefore contribute to the protection of such structures in the dry state. Since the peroxisome was the only major organelle for which we could not identify the targeting of a LEA protein, this suggests that the organelle could be less susceptible to stress-induced damage or that such damage would not have immediate consequences for cell survival. Peroxisomes do not synthesize phospholipids, and although biogenesis of the organelle is enigmatic, the peroxisomal membrane is expected to derive from the ER (Heiland and Erdmann, 2005). It is also postulated that de novo biogenesis from proto-peroxisomes could be a fast process (Lazarow, 2003). It is thus possible that damaged peroxisomes would be rapidly recycled by the pexophagosome, in which LEA9 protein was identified, to allow either elimination of the damaged organelle or possible repair of their membrane by interaction with ER membrane.

Systematic Occurrence of Class A $\alpha$-Helix Motif in the LEA_4 Family

As explained above, LEA42 and LEA48 are putative orthologs of LEAM, a pea mitochondrial protein that was previously shown to fold during dehydration into a peculiar amphipathic helical form, the so-called class A $\alpha$-helix, allowing the protein to immerse laterally within the inner layer of the inner membrane, reinforcing the membrane in the dry state (Grelet et al., 2005; Tolleter et al., 2007, 2010). In the class A $\alpha$-helix motif, negatively charged residues form
a stripe along the axis of the helix, in the middle of the hydrophilic face, which is flanked on each side by two stripes of positively charged residues (Supplemental Figure 17). This allows the helix to interact laterally with membranes by exposing the hydrophobic face of the helix toward the fatty acid core of the membrane, while positively charged residues interact with phosphate groups of phospholipids, allowing the negatively charged residues to interact with polar heads of positively charged phospholipids (Tolleter et al., 2010). Expectedly, the class A α-helix motif found in pea LEAM protein was clearly detected in their putative homologs LEA42 and LEA48, but also in the case of several other LEA proteins located in different compartments (cytosol, ER, plastid, mitochondria, and pexophagosome), as shown in Supplemental Figure 17.

We decided therefore to scan all Arabidopsis LEA proteins for the occurrence of class A α-helix motifs. A computational analysis was performed using the screening module of the HeliQuest Web server (Gautier et al., 2008) that was adjusted in order to detect protein segments with an α-helix structure sharing sequence parameters and physico-chemical features typical of the LEAM protein class A α-helix motifs (see Methods for details). The results showed that such a structural feature was only shared by proteins from the LEA-4 family (Figure 10). Among the 18 proteins in the family, only two of them (LEA23 and LEA24) displayed amphipathic α-helix motifs resembling those of the class A type, but not canonical, and thus considered here as class A-like (Supplemental Figure 18). In particular, the distribution of charges is not as regular as in the case of other proteins of the LEA-4 family. Interestingly, these two paralogous proteins were included in the LEA-4 group by Hundertmark and Hincha (2008) and in our study because they cluster with other LEA-4 proteins, although they do not match significantly with a PFAM domain (e-value of 1.2 for LEA23; no match for LEA24). It is therefore not surprising that these two proteins do not share structural properties such as the class A α-helix motif with true members of the LEA-4 family. We propose that all LEA-4 proteins exhibiting the class A (and class A-like) α-helix motif would interact and protect various cellular membranes during dehydration, as shown for LEAM expressed in pea seed mitochondria (Tolleter et al., 2007, 2010). This hypothesis is supported by the fact that LEA-4 proteins are distributed in many subcellular compartments, which would allow a variety of cellular membranes to benefit from protection by LEA proteins during dehydration resulting from desiccation or freezing. LEA7 was indeed shown to interact in vitro with synthetic membranes in the dry state (Popova et al., 2011). Since LEA7 was found to be a cytosolic protein, it could be involved, like the other cytosolic LEA-4 proteins, in the protection of the plasma membrane or organelar membranes, acting on the cytosol exposed side of the membranes. LEA-4 proteins targeted to organelles would provide protection to the corresponding membranes (e.g., inner mitochondrial membrane, inner membrane of the plastid envelope, thylakoid membrane, ER membrane, tonoplast, and pexophagosome). In the case of plastids, it is questionable whether the interactions of class A α-helix motif would be as favorable as in the case of other cell membranes due to the high proportion of galactolipids, which could not easily interact with the positive charges of the proteins because they lack the necessary phosphate group. Indeed, the lipid composition of the envelope inner membrane of chloroplasts, etioplasts, or proplastids comprises only 10% phospholipid and that of thylakoids is even less, at 8%. Whether this would be sufficient to anchor LEA-4 proteins during dehydration is not known. It is interesting that the two plastid proteins with class A-like motifs (LEA23 and LEA24) were shown to protect membranes in vitro (Thalhammer et al., 2010). Interactions of proteins through such amphipathic α-helix were considered biologically relevant because they were favored by lipid compositions mimicking plastid membranes (Thalhammer et al., 2010). Therefore, plastids could harbor a combination of LEA proteins with either class A (LEA11, 12, 42, and 48) or class A-like (LEA23, 24) α-helix motif, which could protect the galactolipid-enriched plastid membranes.

Interestingly, in a recent analysis of the degradation of metacaspase 9 (a distant relative of the metazoan caspases) in 2-d-old Arabidopsis seedlings, seven LEA-4 proteins were identified as potential substrates of the processing enzyme (Tsiatsiani et al., 2013). Among these, four proteins (LEA25, 28, 29, and 36) are cytosolic, while two display dual location in mitochondria and plastids (LEA42 and LEA48) and one resides in the ER (LEA30). Since the metacaspase is located in the cytosol (Tsiatsiani et al., 2013), this suggests that the three organelar proteins would be processed before import or in the course of organelle and protein turnover since LEA proteins progressively disappear during early germination and growth. Finally, among the eight PFAM LEA families, the LEA-4 family is the most widely distributed, with representatives in plants, metazoans, fungi, heterolobosea, stramenopiles, and bacteria, and it is possible that most of these proteins harbor the class A α-helix motif and thus could play a role in stress protection of biological membranes.

Overall, this analysis of the full complement of LEA proteins within a eukaryotic organism reveals their wide subcellular distribution, with members identified in almost every cellular compartment. Given the current view that all LEA proteins function in stress tolerance, it follows that each compartment requires the presence of LEA protein in certain situations, e.g., dehydration, freezing, cold, or oxidative stress. The predominant cytosolic location agrees with the central position of this compartment, which interacts physically and functionally with all organelles. Cytosolic LEA proteins could therefore be involved in stress protection not only within the cytosol itself, but at the level of membranes delimiting the organelles. The only major organelle for which we could not identify the presence of a LEA protein was the peroxisome, which suggests their rapid biogenesis and recycling through pexophagy could compensate for stress-induced damage.

The complexity of the LEA proteome in plants, which arose from whole-genome duplication and endoreplication events, has been retained along evolution, and the apparent redundancy of LEA proteins in most subcellular compartments likely reflects the versatility of their function in plant life.

METHODS

Plant Culture, Transformation, and Subcellular Localization of Fluorescent Proteins

Arabidopsis thaliana plants (Columbia-0 ecotype) or transgenic lines were grown in Jiffy7 pots (Jiffy Products International) in a growth chamber
Transgenic lines expressing ER-GFP or Perox-CFP were obtained from the ABRC (Nelson et al., 2007). Arabidopsis mesophyll protoplasts were isolated according to a procedure adapted from Yoo et al. (2007) as described by Candat et al. (2013). BFA assays were performed after overnight incubation of transformed protoplasts. Protoplasts were treated with BFA at 10 μg/ml for 1.5 h in the dark prior to observations. Transient Arabidopsis seedling transformation was performed according to the procedure from Marion et al. (2008), with some modifications. Sterilized Arabidopsis seeds were sown on MS 1% agar medium (Murashige and Skoog basal medium, Sigma-Aldrich; Agar HP696 CIV, Kalys) in six-well culture plates (CytoOne reference CC7672-7506) and stratified for 2 d at 4°C. Seedlings were grown for 4 d (same growth conditions as above) prior to transformation. After overnight growth at 28°C, Agrobacterium tumefaciens cells (CS5 strain) carrying appropriate expression vectors were collected and resuspended at OD600 = 1 in 2 mL of Murashige and Skoog medium supplemented with 200 μM acetoxyryngone and 0.05% Tween 20 (Sigma-Aldrich). Seedlings were covered by the Agrobacterium suspension for 30 min in the dark. Excess medium was subsequently removed and the plates were transferred to a growth room for 3 d (same growth conditions as above). For a selection of 11 LEA proteins (LEA9, 13, 21, 22, 30, 37, 38, 41, 42, 43, and 48), as well as for the mito-mCherry, transgenic lines expressing fusion proteins were also established by Agrobacterium floral dip transformation (Clough and Bent, 1998) of wild type or lines already expressing organellar markers for ER (YFP-HDEL; Teh and Moore, 2007), peroxisomes (YFP-SKL; Mathur et al., 2002), or mitochondria (mito-mRFP1). Plants from 3 to 21 independent transgenic lines at the T1 stage were selected for microscopy. Subcellular localization of FP-tagged proteins was examined in protoplasts, cotyledons, or leaves using a Nikon A1 laser scanning confocal microscope and NIS-element software (Nikon). GFP, RFP, and chlorophyll were excited with 488- 561-, or 638-nm laser lines, respectively, with an emission band of 500 to 550 nm for GFP detection, 570 to 620 nm for RFP detection, and 662 to 737 nm for chlorophyll autofluorescence. Experiments using CFP and YFP constructs were performed using the same settings as above for GFP and those with mCherry with the settings for RFP detection. For protoplasts, an average of 50 protoplasts were observed for each experiment (two to five independent protoplast isolation and transformation), with a transformation rate averaging to 25%. In all cases, the subcellular pattern of fluorescent protein was homogeneous, and the images shown in the figures are representative of average protoplasts from a total of 32 transformed protoplasts. Quantitative data are available in Supplemental Table 2 for LEA-FP constructs. Analysis of GFP Fusion Proteins

The purification and the analysis of GFP-tagged proteins were performed as described (Candat et al., 2013). The samples were separated by SDS-PAGE on a 12% acrylamide gel. They were then electroblotted onto immobilon polyvinylidene fluoride membrane (Millipore Merck) at 100 V for 1 h, in 10 mM CAPS, pH 7.4, and 150 mM NaCl containing 1.5% (v/v) Tween 20. The membrane was then incubated for 1 h with a monoclonal anti-GFP antibody conjugated to horseradish peroxidase (Miltenyi Biotec) diluted 1:5000 in TBS buffer containing 0.05% (v/v) Tween 20. The membrane was then washed three times for 10 min with the same buffer. Immunodetection was performed by incubating the membrane for 1 min in 100 mM Tris, pH 8.5, containing 0.4 mM para-hydroxy coumarin acid, 2.5 mM luminol, and 5.43 mM H2O2. Chemiluminescence was monitored by a molecular imager Chemidoc TM XRS system (Bio-Rad). Bioinformatics Analyses

Subcellular localizations were predicted by the following freely available online programs: PSORT (http://psort.hgc.jp/form.html; Nakai and Horton, 1999), MultiLoc (http://abi.uni-tuebingen.de/Services/MultiLoc; Höglund et al., 2006), YLoc (http://abi.uni-tuebingen.de/Services/ YLoc/webloc.cgi; Briesemeister et al., 2010), SubLoc v1.0 (Hua and Hawkins, 2005), and Mitopred (Guda et al., 2004). Default settings were used and plant proteins data sets were selected when available. Predictions were represented as heat map using the Heatmap Builder 1.0 program (King et al., 2005; http://ashleylab. stanford.edu/tools/tools-scripts.html). Helical wheel analyses were performed using the screening module of HeliQuest Web server (Gautier et al., 2000). This module allows the user to screen databases in order to find sequences that have the general physico-chemical features of a target α-helix sequence. We used a personal LEA protein database comprising the 51 Arabidopsis LEA proteins, the mitochondrial pea Pisum sativum) LEAM protein (accession CAF23207), and its ortholog from Artemisia franciscana (accession ACM16888). The module was modified to incorporate new parameters specific to LEAM protein that exhibits class A
α-helix motifs (Tolleter et al., 2007). We used a sliding 36-residue window to scan each of the protein sequence for the presence of segments exhibiting class A α-helix motifs. A protein was selected if at least one of its segments fulfilled a set of sequence parameters and physico-chemical criteria comparable to the class A α-helix segments in LEAM protein. Segments containing at least one cysteine (possibly engaged in a disulfide bridge) and/or a proline (a helix-breaker) were excluded. The mean hydrophobicity (H) and the mean hydrophobic moment (μH) (Eisenberg et al., 1985), the sum (NSTG) of serine, threonine, and glycine residues, the minimum number (Npol) of charged polar residues (Ser, Thr, Asn, and Gln), the minimum number of alanine (NA), the number (Ncha) of charged residues at pH 7.4 (Asp, Glu, Lys, and Arg), and the net charge (z) of the sequence must be within the following ranges: 0.4 ≤ H ≤ 0.05; 0.2 ≤ μH ≤ 0.4; Npol ≥ 2; Ncha ≥ 23; NA ≥ 3; 10 ≤ NSTG ≤ 20; −2 ≥ z ≥ 1. The HeliQuest server generated helical projections for each α-helix segment that were visually examined to validate the class A α-helix predictions.

Phylogenetic Analysis

A rectangular cladogram of the 51 protein sequences was obtained using the Dendroscope 3.1.0 software (Huson and Scornavacca, 2012) and a multiple alignment generated with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The alignment (Supplemental Data Set 1) was generated using Gonnet protein weight matrix (open gap = 10, gap extension = 0.20, gap distances = 5, no end gaps, no iteration, neighbor-joining clustering).

Accession Numbers

Accession numbers from this article can be found in Supplemental Table 3.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Heat Map of Subcellular Localization Predictions of the 51 Arabidopsis LEA Proteins.

Supplemental Figure 2. Transient Expression of Cytosolic LEA Protein in Arabidopsis Protoplasts.

Supplemental Figure 3. Diffusion of Cytosolic LEA-GFP Proteins in the Nucleus.

Supplemental Figure 4. Protein Gel Blot Analysis of LEA Protein Expressed in Protoplasts.

Supplemental Figure 5. Cytosolic Localization of GFP-LEA Fusions for Proteins Localized in Mitochondria or Plastid When Expressed as LEA-GFP Fusions (See Figures 3 and 4).

Supplemental Figure 6. Leaf Cells from Arabidopsis Transgenic Lines Expressing LEA Proteins Targeted to Mitochondria or Dual Targeted to Mitochondria and plastid.

Supplemental Figure 7. LEA13-RFP Does Not Co localize with Golgi in Arabidopsis Protoplasts.

Supplemental Figure 8. Arabidopsis Leaf Cells from Transgenic Lines Expressing LEA Proteins Targeted to the Secretory Pathway or the Pexophagosome.

Supplemental Figure 9. Transient Expression in Arabidopsis Protoplasts of GFP-LEA Fusions for Proteins Which Are Targeted to the Secretory Pathway as LEA-GFP Fusions.

Supplemental Figure 10. Vacuolar and Golgi Localization of LEA 21 in Arabidopsis Protoplasts.

Supplemental Figure 11. Localization of LEA21-GFP and LEA22-GFP in Arabidopsis Protoplasts.

Supplemental Figure 12. LEA11 and LEA12 Sequence Alignment.

Supplemental Figure 13. Estimation of Prediction Software Efficiency.

Supplemental Figure 14. Cytosolic Localization of LEA7 and LEA29 in Arabidopsis Protoplasts.

Supplemental Figure 15. Proteomic Quantification of LEA Proteins in Different Organs of Arabidopsis.

Supplemental Figure 16. LEA2 and LEA38 Sequence Alignment.

Supplemental Figure 17. Modeling of Class A α-Helix Motif of LEA Proteins.

Supplemental Figure 18. Modeling of Class A-Like α-Helix Motif of LEA23 and LEA24.

Supplemental Table 1. Theoretical and Apparent Molecular Masses of LEA and Fusion Proteins.

Supplemental Table 2. Quantitative Data about transiently Transformed Protoplasts.

Supplemental Table 3. Arabidopsis LEA Templates Used for PCR Amplification.

Supplemental Table 4. Primers Used for the Construction of LEA Gene Fusions and Markers.

Supplemental Data Set 1. ClustalW2 Alignment of the 51 LEA Proteins.

ACKNOWLEDGMENTS

This research was supported by a grant from ANR Blanc MITOZEN (ANR-12-BSV8-0021-01). We thank Pauline Poupard (IRHS Angers) and Jean-Pierre Andrieu (IBS, Grenoble) for N-terminal sequence determination and Marjorie Juchaux and Fabienne Simonneau (IMAC QUASAV, Angers) for assistance with the confocal microscope.

AUTHOR CONTRIBUTIONS


Received April 30, 2014; revised June 5, 2014; accepted June 13, 2014; published July 8, 2014.

REFERENCES


The Ubiquitous Distribution of Late Embryogenesis Abundant Proteins across Cell Compartments in Arabidopsis Offers Tailored Protection against Abiotic Stress
Adrien Candat, Gaël Paszkiewicz, Martine Neveu, Romain Gautier, David C. Logan, Marie-Hélène Avelange-Macherel and David Macherel

Plant Cell; originally published online July 8, 2014;
DOI 10.1105/tpc.114.127316

This information is current as of July 7, 2017

Supplemental Data /content/suppl/2014/06/17/tpc.114.127316.DC1.html
eTOCs Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY