Refining the Definition of Plant Mitochondrial Presequences through Analysis of Sorting Signals, N-Terminal Modifications, and Cleavage Motifs

Shaobai Huang, Nicolas L. Taylor, James Whelan, and A. Harvey Millar*

Australian Research Council Centre of Excellence in Plant Energy Biology, M316, University of Western Australia, Crawley, Western Australia 6009, Australia

Mitochondrial protein import is a complex multistep process from synthesis of proteins in the cytosol, recognition by receptors on the organelle surface, to translocation across one or both mitochondrial membranes and assembly after removal of the targeting signal, referred to as a presequence. In plants, import has to further discriminate between mitochondria and chloroplasts. In this study, we determined the precise cleavage sites in the presequences for Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*) mitochondrial proteins using mass spectrometry by comparing the precursor sequences with experimental evidence of the amino-terminal peptide from mature proteins. We validated this method by assessments of false-positive rates and comparisons with previous available data using Edman degradation. In total, the cleavable presequences of 62 proteins from Arabidopsis and 52 proteins from rice mitochondria were determined. None of these proteins contained amino-terminal acetylation, in contrast to recent findings for chloroplast stromal proteins. Furthermore, the classical matrix glutamate dehydrogenase was detected with intact and amino-terminal acetylated sequences, indicating that it is imported into mitochondria without a cleavable targeting signal. Arabidopsis and rice mitochondrial presequences had similar isoelectric points, hydrophobicity, and the predicted ability to form an amphiphilic α-helix at the amino-terminal region of the presequence, but variations in length, amino acid composition, and cleavage motifs for mitochondrial processing peptidase were observed. A combination of lower hydrophobicity and start point of the amino-terminal α-helix in mitochondrial presequences in both Arabidopsis and rice distinguished them (98%) from Arabidopsis chloroplast stroma transit peptides. Both Arabidopsis and rice mitochondrial cleavage sites could be grouped into three classes, with conserved −3R (class II) and −2R (class I) or without any conserved (class III) arginines. Class II was dominant in both Arabidopsis and rice (55%–58%), but in rice sequences there was much less frequently a phenylalanine (F) in the −1 position of the cleavage site than in Arabidopsis sequences. Our data also suggest a novel cleavage motif of (F/Y)1(S/A) in plant class III sequences.

Plant mitochondria play a key role in energy production and metabolism that requires the import and assembly of at least 1,000 proteins. Protein import into mitochondria begins with synthesis of the precursor protein in the cytosol, followed by binding to various proteins in the cytosol, binding to receptors on the outer mitochondrial membrane, translocation across one or both mitochondrial membranes, removal of the targeting signal, termed a presequence, and intraorganelar sorting and assembly. A variety of studies have shown that there is no primary amino acid sequence conservation among presequences, but they do have a high proportion of positively charged residues and the capacity to form an amphiphilic α-helix (Roise et al., 1986; von Heijne, 1986). Many mitochondrial presequences have a loosely conserved motif near the cleavage site comprising an Arg residue at the −2 and/or −3 position (von Heijne et al., 1989; Schneider et al., 1998). This Arg has been experimentally shown to be an important recognition site for the mitochondrial processing peptidase (MPP; Arretz et al., 1994; Ogishima et al., 1995; Tanudji et al., 1999). MPP is a heterodimeric enzyme that contains two similar subunits: α-MPP is involved in binding precursor proteins and β-MPP catalyzes the cleavage of the presequence (Kitada et al., 1995; Luciano et al., 1997). In yeast and mammals, MPP is a soluble protein located in the matrix, but in plants, MPP is integrated into the inner membrane-bound cytochrome b/c1 complex (Braun et al., 1992; Eriksson et al., 1994; Glaser and Dessi, 1999).

The mechanism through which the targeting signal binds to a receptor protein has been revealed by NMR studies and the crystal structure of rat Tom20 (for translocase of the outer membrane) with a bound presequence (Abe et al., 2000; Saitoh et al., 2007). A dynamic binding model in which different hydrophobic residues in the presequence interact with Tom20...
Plant Mitochondrial Presequences

has been proposed. Thus, the presequence has mobility in the binding site via hydrophobic interactions, with several different binding states being possible. This model accounts for the ability of a single Tom20 in yeast to bind to a diverse array of presequences. Although plants contain a protein that is called Tom20 and that has a receptor function in mitochondrial import, it is not orthologous to yeast or mammalian Tom20 (Perry et al., 2006; Lister et al., 2007). However, the NMR structure of plant Tom20 reveals a similar hydrophobic binding pocket. This has been highlighted as a case of convergent evolution of a receptor that uses a similar mechanism of binding to recognize presequences (Lister and Whelan, 2006). Although structural studies reveal the importance of hydrophobic residues for presequence binding, several studies on yeast, mammals, and plants reveal an important role for positively charged residues in presequences for import into mitochondria (Lister et al., 2005; Neupert and Herrmann, 2007). These positively charged residues may play a role in positioning the amphipathic α-helix for binding to Tom20 and also in subsequent translocation into and across the pores forming proteins of the TOM and TIM (for translocase of the inner membrane) complexes. Movement of the presequence into and across a translocase is explained by the binding chain hypothesis (Pfanner and Geissler, 2001). According to this hypothesis, a presequence binds to higher affinity sites in the import apparatus until it is “trapped” on the inside of the inner membrane by a combination of electrostatic interactions, the net negative charge on the inside of the inner membrane, and binding to matrix-located HSP70 (Zhang and Glaser, 2002).

In addition to the fact that plant Tom20s are not orthologous to other Tom20s, plant mitochondria also lack the other two receptor components that have been functionally characterized in yeast, namely Tom70 and Tom22 (Lister et al., 2007). Furthermore, mitochondrial and plastid targeting signals contain significant similarities in plants; thus, plant mitochondrial presequences have evolved to differentiate from the large number and abundant nature of plastid proteins requiring import from the cytosol (Macasev et al., 2000). This raises the question of how similar plant mitochondrial targeting signals are to those of yeast and how they are differentiated from plastid transit peptides. To adequately address these questions, a large number of presequences need to be assembled to define motifs that differentiate presequence classes. Traditionally, the N-terminal sequences of plant mitochondrial proteins have been obtained by Edman degradation either from purified mitochondrial protein complexes or in proteome studies (Braun and Schmitz, 1995; Jänisch et al., 1996; Millar et al., 1998, 1999; Kruft et al., 2001; Bardel et al., 2002). The presequences could only be obtained by comparison of these N-terminal sequences with the preprotein sequence deduced from full-length cDNA sequences, which were only available in a small number of cases. Glaser et al. (1998) presented a list of approximately 100 plant mitochondrial presequences; these were mainly derived from prediction and/or comparisons in homologous cDNA-derived protein sequences with a core set of 31 experimentally proven presequences for plant mitochondrial proteins. Later analysis of 58 experimentally proven plant mitochondrial presequences deposited in the Swiss-Prot database revealed two major classes containing an Arg residue at positions −2 and −3 and one class without any conserved Arg residues (Zhang et al., 2001; Zhang and Glaser, 2002). However, this data set relied on the sequences available at the time that were from a variety of plant species and contained redundant orthologs from similar proteins. This data set also clearly focused on dicot plants, as less than 20% of the sequences were from monocot species.

In the chloroplast, N-terminal modification of chloroplast proteins has been shown to be important for protein viability (Pesaresi et al., 2003). N-terminal acetylation can be detected by high-resolution mass spectrometry (MS) through a change in mass of the N-terminal peptide. The recent systematic analysis of the Arabidopsis (Arabidopsis thaliana) chloroplast proteome revealed 47 stroma proteins with N-acetylated residues and 62 without N-acetylated residues (Zybailov et al., 2008). The detection of N-terminal and non-N-terminal acetylated proteins by identifications of semitryptic peptides also allowed analysis of the cleavage sites and potential motifs for cleavage by processing peptidases (Zybailov et al., 2008). However, no systematic experimental analysis of N-terminal modifications and potential cleavage sites of plant mitochondrial proteins has been carried out to date using such an MS approach.

In this study, we have determined Arabidopsis and rice (Oryza sativa) mitochondrial protein-targeting presequences and cleavage sites using an MS approach after gel- or liquid chromatography (LC)-based separation and also identified a range of N-terminal modifications of mitochondrial proteins. Validation of this method was performed by false-positive analysis and comparison with previous results in Arabidopsis using an Edman degradation approach (Kruft et al., 2001). We compared the characteristics of the generated Arabidopsis and rice mitochondrial presequences and the cleavage site motifs. Comparison with experimentally proven yeast mitochondrial presequences and Arabidopsis plastid stroma transit peptides allowed consideration of some evolutionary questions and insights into the different signal-recognizing mechanism(s) used to distinguish between organelles.

RESULTS

Validation of MS-Identified Semitryptic Peptides for Defining the Putative N Terminus of Mature Mitochondrial Proteins

We separated Arabidopsis and rice mitochondrial proteins using gel-based (isoelectric focusing [IEF]-
SDS-PAGE) methods and digested the protein spots with trypsin. We also performed non-gel-based LC separation of peptides derived from trypsin digestion of whole Arabidopsis and rice mitochondria. In conventional protein identifications after trypsin digestion using MS/MS peptide spectra, searches are usually undertaken using parameters that require peptides to be products of a defined protease digestion (e.g. trypsin digestion) of a set of precursor protein sequences predicted from gene models. Such analysis will not identify spectra for the N-terminal peptides of mature mitochondrial proteins that are derived from the processing carried out by MPP. In this study, we have exploited our lists of the Arabidopsis and rice mitochondrial proteomes (Heazlewood et al., 2004, 2007; Huang et al., 2009) to build smaller databases in order to carry out searches of archived MS/MS spectra with the digestion parameter set to “no enzyme.” This allows matching of any peptides after trypsin digestion, even those derived from digestion with another enzyme such as MPP.

Due to the wide latitude in the search criteria, this analysis yielded many thousands of spectral matches. Using the mitochondrial database and following selection of the most N-terminal peptide matches for each protein with an ion score greater than 20 and $P < 0.05$, the false-positive detection rates for rice and Arabidopsis matches by decoy were 4.5% and 3.6%, respectively, which were calculated using an equation described previously (Elias et al., 2005). To further interrogate this set, and to deduce the validity of semitryptic peptide matches, we classified these peptides into three groups. First, peptides that were derived at both N and C termini by cleavage at a non-tryptic site. Second, peptides that were derived by a C terminus consistent with trypsin cutting but an N-terminal cleavage at a non-tryptic site. Third, peptides that were derived at both N and C termini by cleavage at a trypsin-cutting site. Displaying these groups according to ion score revealed that the first group dominated the set of spectra with ion scores of 20 to 24 but rapidly decreased as a proportion of the matched spectra as the ion score was raised to greater than 37 (Fig. 1). We concluded that this group contained biologically unlikely matches, as the samples were digested with trypsin and thus random peptide cleavages were unlikely. In contrast, the second group of semitryptic peptides were distributed across the four ion score ranges but peaked in the $\geq 37$ ion score group, while the third tryptic peptide group was more biased to the $\geq 37$ ion score group (Fig. 1). In rice, if the first group is considered a refined measure of false-positive matches, then this decreases from being approximately 50% with ion scores of 20 to 24 to being less than 1.4% for peptides with ion scores above 37 (Fig. 1A). A similar pattern was also observed using typical spectral sets from Arabidopsis (Fig. 1B). On this basis, semitryptic peptides with ion scores greater than 37 were selected for further analysis as high-quality spectral matches to the mitochondrial proteome and as putative presequence cleavage sites. The false-detection rates for these N-terminal peptides with an ion score threshold of 37, using a decoy random database, were 0.93% for rice and 1.09% for Arabidopsis.

To confirm that our method was not overly influenced by the use of the restricted mitochondrial protein database, we searched a range of the gel spot files against the Arabidopsis proteome (The Arabidopsis Information Resource 8) and the rice proteome (The Institute for Genomic Research osa5). There was no difference in the total number of identified peptides for each protein, the ion scores were unchanged, and in each case the N-terminal peptide was still the top match to the protein identified (data not shown).

A further validation of these N-terminal semitryptic peptides was conducted by comparison with the limited number of experimentally proven cleavage sites of known Arabidopsis mitochondrial proteins from the literature (Kruft et al., 2001; Supplemental Table S1). Seventeen Arabidopsis mitochondrial precursors were identified by N-terminal Edman degradation of gel-separated mitochondrial proteins (Kruft et al., 2001); semitryptic peptides for eight of the same pro-
Proteins were also found in our study, and all eight showed identical cleavage sites (Supplemental Table S1). In total, we identified N-terminal peptides for 80 Arabidopsis mitochondrial proteins and 77 rice mitochondrial proteins. Compared with the Edman degradation method, the MS approach allows deeper analysis of more mitochondrial proteins in the absence of gel-based separations.

Proteins with Noncleavable N-Terminal Sequences

Among the N-terminal semitryptic peptides, 25 in rice and 18 in Arabidopsis indicated the absence of a presequence, as the peptide matched very near to the beginning of the predicted precursor protein with or without removal of Met (Table I). Removal of Met by Met aminopeptidase cleavage is known to depend on the length of the side chain of the second amino acid (Sherman et al., 1985; Giglione and Meinnel, 2001). If the second incorporated amino acid is Ala (A), Gly (G), Pro (P), Ser (S), or Thr (T), the initiator Met is systematically removed. If the second side chain is large [such as Arg (R), Asn (N), Asp (D), Glu (E), Ile (I), Leu (L), or Lys (K)], the initiator Met is retained at the N terminus of the mature proteins, while Val (V) is intermediate for the specificity of Met aminopeptidase, resulting in both cleaved and uncleaved Met followed by Val. As shown in Table I, this specificity is evident in the identified mitochondrial protein N termini from both Arabidopsis and rice. The Met is retained only when the second amino acid residue is E, L, F, D, or N (Table I). This observation also represents another level of independent validation of the identified semitryptic peptides derived from the MS data.

Many of the proteins detected without a cleaved N terminus from rice and Arabidopsis were mitochondrion outer membrane or intermembrane space located, such as porins, OMP85, small TIMs, and adenylate kinases. A further set were inner membrane proteins that may not enter the matrix but directly insert into the inner membrane from the intermembrane space (e.g. UCP3, COX VIIa, and complex I and III subunits; Table I). However, in both rice and Arabidopsis, a notable exception was Glu dehydrogenase (GDH), which had an N-terminal acetylated intact sequence (Table I; Supplemental Fig. S1). Specific peptides from other regions of the different GDH isoforms listed in Table I, from gel protein spots, confirm that the identical N-terminal peptide can be found from each isoform. In Arabidopsis, 12 of the 16 nonprocessed proteins had their N-terminal residue acetylated (Table I), the only exceptions being the outer membrane porins and an unknown function protein (Table I). In rice, the frequency of the N-terminal acetylation was lower (only 12 of 22), with a range of respiratory chain components, unknown function proteins, and carrier proteins lacking acetylation. Conservation of some proteins, such as adenylate kinases and prohibitin, with N-terminal acetylation in both species was also observed (Table I). For proteins with noncleavable sequences detected only by LC, we cannot rule out the possibility that removed presequences are stable and have been detected by MS. But presequence turnover is considered to be very rapid in mitochondria due to the activity of PreP (Moberg et al., 2003), so we consider this unlikely.

Proteins with Presequence Processing

In total, the N-terminal peptides of 62 proteins from Arabidopsis and 52 proteins from rice indicated the cleavage of N-terminal presequences (Supplemental Table S2). None of these 114 semitryptic peptides detected from Arabidopsis and rice mitochondrial proteins had N-terminal acetylation, which contrasts to the observation of many proteins in the chloroplast stroma with N-terminal acetylation after presequence cleavage (Zybailov et al., 2008). It appears that the plant mitochondrial matrix may lack the necessary machinery to catalyze N-terminal acetylation. Consistent with this, the N terminus of mitochondrial genome-encoded protein ATP synthase Fδ subunit 1 (Osmtg00581) was not acetylated (Table I), while the N-terminal acetylated GDHs (Table I) are likely to have been modified before import into the mitochondrial matrix.

Characteristics of Presequences in Mitochondrial Proteins

Comparisons between sequences of N-terminal semitryptic peptides with the amino acid sequences deduced from the corresponding genes enabled us to define the cleaved presequences (Supplemental Table S2). We compared them with 105 experimentally proven yeast mitochondrial protein presequences (Supplemental Table S3) and 47 Arabidopsis chloroplast stroma transit peptides (Supplemental Table S4; Zybailov et al., 2008). The length of mitochondrial presequence varied greatly from 19 to 109 amino acids in Arabidopsis, 18 to 117 amino acids in rice, and six to 122 amino acids in yeast (Fig. 2A). The majority of these proteins (approximately 74%) had 21 to 70 residues (Fig. 2A). The averaged amino acid lengths of the observed presequences in Arabidopsis and rice were 50 and 45, respectively, compared with a significantly shorter average length of only 31 residues in yeast (Fig. 2A; Table II). The plastid stroma transit peptides were typically longer (Table II), with most between 50 and 70 amino acids (Fig. 2A).

The amino acid composition of the presequences between Arabidopsis and rice differed. Most notably, the Ala composition in rice (19.8%) was double that in Arabidopsis (9.7%) presequences, and the Ser composition in rice (11.4%) was lower than that in Arabidopsis (16.2%; Fig. 2B). The four dominant amino acid residues (S, R, A, and L) represented 47% and 57% of the total residues in the presequences of Arabidopsis and rice, respectively (Fig. 2B). Compared with Arabidopsis, rice also had a higher pro-
portion of the amino acid residues G and P but had lower or similar compositions of the rest of the amino acid residues (Fig. 2B). Overall, it appeared that yeast and Arabidopsis mitochondrial presequences were more similar to each other in amino acid composition than either one was to the rice presequence set (Fig. 2B).

Interestingly, the Arabidopsis stroma transit peptides had no consistent differences in amino acid composition when compared with the mitochondrial presequences (Fig. 2B). More detailed analysis of the first 10 amino acids following Met showed a clear decreased proportion of Arg (R) and increased Pro (P) in

Table I. Rice and Arabidopsis mature mitochondrial proteins with uncleavable presequences with or without removal of the first Met

<table>
<thead>
<tr>
<th>GI</th>
<th>Description</th>
<th>TP</th>
<th>Met</th>
<th>Semitypic Peptide</th>
<th>Score</th>
<th>Expt</th>
<th>Acetyl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Os03g42799</td>
<td>Expressed protein</td>
<td>35</td>
<td>–</td>
<td>MFLGAI PR</td>
<td>40</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os08g33460</td>
<td>Expressed protein</td>
<td>–</td>
<td>–</td>
<td>MNTDITASVKPEPVYDR</td>
<td>116</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os03g03770</td>
<td>Complex I B18-like subunit</td>
<td>–</td>
<td>–</td>
<td>MEAAAAVQGLTSKPOATQAEMSEAR</td>
<td>81</td>
<td>Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os08g42380</td>
<td>TIM8</td>
<td>–</td>
<td>–</td>
<td>MENSAEMQQR</td>
<td>51</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os07g10500</td>
<td>Complex III 14-kD protein-like</td>
<td>78</td>
<td>–</td>
<td>MLSSLSAWLVPVR</td>
<td>69</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os12g13380</td>
<td>Adenylate kinase A</td>
<td>–</td>
<td>M</td>
<td>AANLEDVPSMELMTELLR</td>
<td>118</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os11g20790</td>
<td>Adenylate kinase B</td>
<td>–</td>
<td>M</td>
<td>AAAANLEDVPSMDLMNELLR</td>
<td>154, 41</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os08g15170</td>
<td>ATP synthase α-chain</td>
<td>13</td>
<td>M</td>
<td>SATSAAAAPFWR</td>
<td>56</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os05g30030</td>
<td>Catalytic/oxidoreductase, acting on NADPH</td>
<td>24</td>
<td>M</td>
<td>AAAAAAEGLAAYR</td>
<td>58</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os07g45320</td>
<td>COX VIIa-like protein</td>
<td>–</td>
<td>M</td>
<td>TEAPFVPR</td>
<td>48</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os03g48110</td>
<td>Expressed protein</td>
<td>10</td>
<td>M</td>
<td>AESPENAAAPAAPAPAPTPAPPPSSPPTK</td>
<td>65</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os04g02670</td>
<td>Expressed protein</td>
<td>22</td>
<td>M</td>
<td>PFTCPYCYSVSTLALVAR</td>
<td>65</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os11g48040</td>
<td>Mitochondrial uncoupling protein 3</td>
<td>–</td>
<td>M</td>
<td>PEHCSKPIDFSGR</td>
<td>68</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os06g50000</td>
<td>Complex I B22-like subunit</td>
<td>75</td>
<td>M</td>
<td>ATSGFLAR</td>
<td>54</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os05g39390</td>
<td>Outer membrane protein (OMP85)</td>
<td>–</td>
<td>M</td>
<td>ATAAAGQPNPDAEHR</td>
<td>93</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os09g19734</td>
<td>Porin</td>
<td>–</td>
<td>M</td>
<td>VGCPLYPEIGK</td>
<td>54, 48</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os05g59500</td>
<td>Porin</td>
<td>–</td>
<td>M</td>
<td>AAAAPPAGCPSYDIGIN</td>
<td>57, 52</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os04g38900</td>
<td>Prohibitin</td>
<td>24</td>
<td>M</td>
<td>AGGPAAVSFLTEINAK</td>
<td>91, 82</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os03g59220</td>
<td>Complex III 14-kD protein-like</td>
<td>81</td>
<td>M</td>
<td>SMLSAFSGWVFR</td>
<td>93</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>Os07g48244</td>
<td>Complex III 6.7-kD protein-like</td>
<td>27</td>
<td>M</td>
<td>ALPASSGFLR</td>
<td>66</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os05g33210</td>
<td>Complex III 8.0-kD protein-like</td>
<td>60</td>
<td>M</td>
<td>GLWDALYR</td>
<td>41</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Matrix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Os1mg00080</td>
<td>ATP synthase F0 subunit 1</td>
<td>–</td>
<td>M</td>
<td>MEFPSR</td>
<td>37</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>Os03g58040</td>
<td>Glu dehydrogenase</td>
<td>18</td>
<td>M</td>
<td>MNALTAASR*</td>
<td>73</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>Os04g45970</td>
<td>Glu dehydrogenase 2</td>
<td>18</td>
<td>M</td>
<td>MNALTAASR*</td>
<td>52</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>AT5G50810</td>
<td>TIM8</td>
<td>–</td>
<td>M</td>
<td>MDPGNPENPELQFLAQKE</td>
<td>74, 75</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>AT3G46560</td>
<td>TIM9</td>
<td>–</td>
<td>M</td>
<td>MDSMMAMGLDGLPEEK</td>
<td>68, 57</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>AT4G23885</td>
<td>Unknown protein</td>
<td>–</td>
<td>M</td>
<td>MNVHVHLK</td>
<td>37</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>AT4G15949</td>
<td>Fumarylacetocatase hydrolase family protein</td>
<td>M</td>
<td>ATSMIQR</td>
<td>60, 44</td>
<td>LC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AT5G50370</td>
<td>Adenylate kinase</td>
<td>M</td>
<td>ATSSAASVDMEIQTVDLMESELLR</td>
<td>108, 91</td>
<td>LC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AT5G63400</td>
<td>Adenylate kinase</td>
<td>M</td>
<td>ATGGAAADLEDVQTVTDLMESELLR</td>
<td>112, 113</td>
<td>LC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AT1G22840</td>
<td>Cytochrome c1</td>
<td>M</td>
<td>ASFDDEAPPGNK</td>
<td>51, 46</td>
<td>LC/Gel</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AT5G40770</td>
<td>ATPH83 (prohibitin 3)</td>
<td>20</td>
<td>M</td>
<td>GSQAAAVSFLSINLAK</td>
<td>65, 54</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>AT3G62810</td>
<td>Complex 1 family protein/LVR family</td>
<td>12</td>
<td>M</td>
<td>VGEALIAYR</td>
<td>86, 79</td>
<td>LC</td>
<td>No</td>
</tr>
<tr>
<td>AT3G12260</td>
<td>Complex I B14 subunit</td>
<td>66</td>
<td>M</td>
<td>AAFPALR</td>
<td>48</td>
<td>Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>AT3G01280</td>
<td>Porin</td>
<td>–</td>
<td>M</td>
<td>VGCPLYTEIGK*</td>
<td>43, 41</td>
<td>LC/Gel</td>
<td>No</td>
</tr>
<tr>
<td>AT5G15090</td>
<td>Porin</td>
<td>–</td>
<td>M</td>
<td>VGCPLYTEIGK</td>
<td>38</td>
<td>LC/Gel</td>
<td>No</td>
</tr>
<tr>
<td>AT5G67500</td>
<td>Porin</td>
<td>–</td>
<td>M</td>
<td>SKGPGLTIDIGK</td>
<td>62, 47</td>
<td>LC/Gel</td>
<td>No</td>
</tr>
<tr>
<td>AT4G20150</td>
<td>Similar to Os07g0498300</td>
<td>69</td>
<td>M</td>
<td>PISATMVGALLGLGTQMYNSALR</td>
<td>41</td>
<td>Gel</td>
<td>No</td>
</tr>
<tr>
<td>AT2G29530</td>
<td>TIM10</td>
<td>–</td>
<td>M</td>
<td>ASPITPVGVTK</td>
<td>63, 61</td>
<td>LC</td>
<td>Yes</td>
</tr>
<tr>
<td>AT1G15120</td>
<td>Complex III 7.8-kD protein</td>
<td>M</td>
<td>ADDEVVDPKK</td>
<td>54, 51</td>
<td>LC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G07440</td>
<td>Glu dehydrogenase (α-subunit)</td>
<td>18</td>
<td>–</td>
<td>MNALATNAR*</td>
<td>60.6</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
<tr>
<td>AT5G18170</td>
<td>Glu dehydrogenase (β-subunit)</td>
<td>18</td>
<td>–</td>
<td>MNALATNAR*</td>
<td>60.9</td>
<td>LC/Gel</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Arabidopsis stroma transit peptides (Fig. 2C), which was consistent with previous findings (Bhushan et al., 2006).

The averaged pI values of the observed presequence peptides in Arabidopsis and rice were basic and very similar (Table II; Fig. 2D). The percentages of prese-
Huang et al.

found that 71% of the presequences were predicted to depending on the length of the mitochondrial prese-
presequences had more than two occur at the beginning of the sequence. Some of the formation of an (Table II). With the possible removal of the first Met, the presequences in rice were predicted to form anmitochondrial presequences were predicted to form
(www-jpred/; Cole et al., 2008). In total, 29% and 11% of amino acid residues in Arabidopsis mitochondrial presequences were predicted to form α-helix (H) and β-sheet (E), while the corresponded
numbers in rice were 34% and 5% (calculated from Supplemental Table S2). In the first 10 amino acids, 52 of 62 (84%) presequences in Arabidopsis and 47 of 52 (90%) presequences in rice were predicted to form an α-helix (Table II). With the possible removal of the first Met, the formation of an α-helix in the presequence is likely to occur at the beginning of the sequence. Some of the presequences had more than two α-helical domains, depending on the length of the mitochondrial presequence (Supplemental Table S2). The formation of β-sheet domain(s) in rice and Arabidopsis mitochondrial presequences was rare (Table II). In yeast, we found that 71% of the presequences were predicted to form an α-helix in the first 10 amino acids, which was lower than those in Arabidopsis and rice, while β-sheet domains were much more common (Table II). The stroma transit peptides had less predictable structure, with only 30% with clear α-helix domains but over 21% with β-sheet domains (Table II).

We also calculated the hydrophobicity values of the first 10 amino acids in presequences using TopPred (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/) and the averaged hydrophobicity values of the first 10 amino acids were −1.30 for Arabidopsis and −1.39 for rice (Table II), and most of them were less than zero (Fig. 2E). In yeast, the average hydrophobicity value of the first 10 amino acid was −1.21 (Table II). These results suggest that mitochondrial presequences had hydrophilic N-terminal sequences for import into mitochondria; in comparison, the stromal transit peptides were substantially less hydrophilic, with an average of only −0.2 (Table II) and a substantial number having positive overall hydrophobicity values (Fig. 2E). Remarkably, only one of the 47 stroma transit peptides in the first 10 amino acids had the combination of the ability to form α-helix and a hydrophobicity value less than the mitochondrial average of −1.30 (Supplemental Table S2). This protein (At1g10760) is a phosphoglucon water dikinase and could be a dually targeted protein due to its identification in Arabidopsis mitochondria (Heazlewood et al., 2004) and plastids (Peltier et al., 2006; Zybailov et al., 2008).

Characteristics of Cleavage Sites of Mitochondrial Proteins

Defining the presequences allowed us to define the peptide cleavage sites for a considerable number of mitochondrial proteins (Supplemental Table S2). A prediction of the cleavage sites of the same mitochondrial proteins was also performed using the most widely used bioinformatic tool, TargetP (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson et al., 2000). In Arabidopsis, 20 of the 62 (32%) sequences had the same cleavage site in the predicted and observed sets, and in rice, the number was 15 of 52 (29%; Supplemental Table S2). Only approximately 30% prediction accuracy of cleavage sites by TargetP suggested the necessity of cleavage site identification using an experimental approach and also a greater need to investigate consensus sequences at cleavage sites in order to improve future prediction.

The 114 mitochondrial presequences represented a diverse and expanded pool for further analysis of the

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Arabidopsis Mitochondria</th>
<th>Rice Mitochondria</th>
<th>Yeast Mitochondria</th>
<th>Arabidopsis Stroma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>49.6</td>
<td>44.7</td>
<td>31.2*</td>
<td>54.2</td>
</tr>
<tr>
<td>pI</td>
<td>11.61</td>
<td>11.84</td>
<td>11.32</td>
<td>11.27</td>
</tr>
<tr>
<td>Hydrophobicity (1–10 amino acids)</td>
<td>-1.30</td>
<td>-1.39</td>
<td>-1.21</td>
<td>-0.20*</td>
</tr>
<tr>
<td>α-Helix (1–10 amino acids)</td>
<td>52/62 (83.9%)</td>
<td>47/52 (90.4%)</td>
<td>75/105 (71.4%)</td>
<td>14/47 (29.8%)*</td>
</tr>
<tr>
<td>β-Sheet (1–10 amino acids)</td>
<td>2/62 (3.2%)</td>
<td>1/52 (1.9%)</td>
<td>19/105 (18.1%)*</td>
<td>10/47 (21.3%)*</td>
</tr>
</tbody>
</table>
characteristics of mitochondrial presequence cleavage sites. We extracted the experimentally proven 57 plant mitochondrial presequences from the Swiss-Prot database (Supplemental Table S5), which was very close to the previously reported number of 58 (Zhang et al., 2001; Zhang and Glaser, 2002). Only 19 Arabidopsis and five rice proteins were listed in the extracted set (Supplemental Table S5). Further literature searches for experimentally proven plant mitochondrial presequences (Supplemental Table S5) added 23 proteins to this list but did not add any new Arabidopsis and rice mitochondrial presequences. Therefore, the current data set provides an enriched set for comparison of cleavage characteristics of the two main model plants, which are also a dicot and a monocot. To identify conserved amino acid residue(s) at cleavage sites, we performed sequence logo analysis (http://weblogo.berkeley.edu/logo.cgi; Crooks et al., 2004) using the 15 C-terminal residues of the presequences and the 15 residues from the N terminus of the mature proteins (Fig. 3). Overall, the 114 Arabidopsis and rice mitochondrial proteins appeared to have a conserved motif of RX(F/Y/L)↓(S/A)(S/T) (where X denotes any amino acid and ↓ denotes the cleavage site; Fig. 3). We did not find in the plant sequences any overrepresentation of Arg at position −10 or the characteristic octapeptide reported in mammalian and yeast mitochondrial presequences that are cleaved first by MPP and then by mitochondrial intermediate peptidase (Isaya et al., 1991). There are some differences in the region of cleavage sites between rice and Arabidopsis, with clear overrepresentation of Phe (F) in the −1 position in Arabidopsis and high frequency of Ala (A) in the mature protein sequence upstream of the cleavage region in rice (Fig. 3). Comparison with the stromal cleavage site showed the distinction of the mitochondrial signature for MPP cleavage versus the plastid stroma processing peptidase cleavage (Fig. 3).

Previously, the plant mitochondrial presequences were classified into three groups based on conserved Arg position: class I (−2R), class II (−3R), and class III (no conserved Arg; Sjoling and Glaser, 1998; Zhang

Figure 3. Consensus sequences of the sites of mitochondrial presequence cleavage and comparison with plastid stroma transit peptide cleavage. Sequence logo analysis of the presequence cleavage site, constructed for Arabidopsis (62) and rice (52) proteins with semi-trypptic peptides derived from MS data analysis, and sequence logo analysis of presequence cleavage sites of 92 yeast mitochondrial presequences and 47 Arabidopsis plastid stroma transit peptides (Zybailov et al., 2008). Sequences used for sequence logo analysis are given in Supplemental Tables S2 to S4. Thirteen yeast mitochondrial proteins with presequences of less than 15 amino acids (Supplemental Table S3) were not included in the sequence logo analysis.
et al., 2001). The class II (-3R) proteins represented 55% and 58% of total Arabidopsis and rice proteins, respectively (Fig. 4). In class II (-3R), sequence logo analysis indicated that Phe (F) in the -1 position in Arabidopsis dominated, while in rice, Y, F, and L were evenly distributed (Fig. 4). In class I (-2R), Arabidopsis sequences appear to have a conserved motif of RXYS (Fig. 4), as was noted previously in an analysis of plant mitochondrial presequences from a variety of species (Zhang et al., 2001). But for rice, such a conserved motif was not conclusive, perhaps due to the limited number of sequences for analysis rather than any real biological difference (Fig. 4). In class III (nonconserved Arg) presequences, both Arabidopsis and rice appear to have a motif of (F/Y)YS (S/A) (Fig. 4). Such an observation is in contrast to the previous report that class III presequences had no significant conservation in amino acid content around the cleavage sites (Zhang et al., 2001), perhaps due to the smaller sample size available at the time. Interestingly, yeast non-R sequences lacked any similar motifs, with an even larger set of 24 sequences in this class. The yeast data set also had other motif(s) based on R in the -10 position, which was likely to be cleaved by the mitochondrial intermediate peptidase (Isaya et al., 1991). This additional group with a motif of RX(Y/L)XX(YS/T) was most similar to the class II (-3R) group in plants but shifted by eight amino acids from the cleavage site.

We assessed the similarity between the presequences of orthologous proteins in the rice and Arabidopsis data sets (Supplemental Table S6). There were 19 sets of cross-species orthologs. There was no apparent conservation of the amino acid sequence of mitochondria presequences and also no clear conservation of cleavage motif in most cases. The most striking similarity was in presequence length, which was well conserved within 17 of the 19 ortholog sets but ranged from 20 to 100 amino acids across the proteins.

We also investigated the predicted secondary structure of 30 amino acids around the cleavage sites (Supplemental Table S2) to test the possible linkage between the cleavage processing and α-helix formation, as suggested previously (Sjoling and Glaser, 1998). We found the numbers of proteins to be predicted to form α-helix and/or β-sheet across the cleavage site to be similar (Supplemental Fig. S2). Furthermore, there was no linkage between the formation of α-helix and β-sheet across the three classes (I–III) of cleavage sites (Supplemental Table S2). All of these results obtained from Arabidopsis and rice suggested that an α-helix is not essential at cleavage sites based on our data set.

**DISCUSSION**

We have determined Arabidopsis and rice mitochondrial protein N-terminal modifications, presequences, and cleavage sites using a MS approach.
based on identification of semitryptic peptides. This approach was validated by false-positive matching predictions and comparison with previous results using an Edman degradation approach (Kruft et al., 2001). These validations were further enhanced by the observation that the peptides matched lacking only the N-terminal Met were nearly all mitochondrial proteins located in membrane or intermembrane space locations and all clearly followed the known specificity of Met aminopeptidase cleavage, as reported previously (Sherman et al., 1985; Giglione and Meinnel, 2001). Compared with Edman degradation and site-directed mutagenesis methods, the MS approach combined with gel and LC separation of peptides/proteins allowed direct and broader analysis of mitochondrial proteins for single amino acid resolution presequence analysis, albeit limited by the fact that the MS cannot identify semitryptic peptides of less than approximately six amino acid residues. In total, a data set of mitochondrial presequences for 62 Arabidopsis proteins and 52 rice proteins was generated, which significantly expands the known experimentally proven plant mitochondrial presequences currently deposited in the Swiss-Prot database and greatly enhances the limited number of mitochondrial presequences known from monocot plants (Supplemental Table S5).

Surprisingly, we found the classic mitochondrial matrix protein, GDH, with an intact and N-terminal acetylated presequence without any cleavage evident in either Arabidopsis or rice (Table II). GDHs function in the catabolism of Glu and in assimilating free ammonia, thus representing an important branch point between carbon and nitrogen metabolism. In mitochondria, GDH is considered to function in the deaminating direction to form 2-oxoglutarate and contribute to NADH production for further ATP generation (Thomas, 1978; Aubert et al., 2001). Interestingly, in human and bovine liver mitochondria, the GDHs P00367 and P00366 contain 53 and 57 amino acid presequences, respectively (Moon and Smith, 1973; Julliard and Smith, 1979; Hochstrasser et al., 1992). Alignment of the mammalian and plant mitochondrial GDHs indicates that the plant N-terminal regions that appear to be responsible for mitochondrial targeting are much shorter than the presequences of the mammalian GDH proteins (data not shown). We recently observed the two Arabidopsis mitochondrial GDHs (At5g07440 [GDH2] and At5g18170 [GDH1]) shifting pI or molecular mass on two-dimensional gels of mitochondrial extracts from shoots and cultured stem cells (Lee et al., 2008a). Reanalysis of peptide spectra from those gel spots has revealed that both had the intact N-terminal sequence with acetylated Met as shown in Table I, indicating that presequence processing is not the case of these observed pI differences (data not shown). The physiological or biochemical significance of the plant matrix GDH lacking a cleavable N-terminal presequence is not clear.

The protein import mechanisms of mitochondria and chloroplasts have a variety of common features, but each is sufficiently different to allow the specificity of proteins to be imported (Macasev et al., 2000; Zhang and Glaser, 2002). This selectivity for import is determined by discrete targeting signals in mitochondrial and chloroplastic precursor proteins and by the different protein complexes located in the outer membrane of organelles that recognize these sequences (TOM and TOC complexes, respectively). Some chloroplastic transit peptides look like mitochondria-targeting peptides and in vitro can be imported into fungal mitochondria (Hurt et al., 1986; Brink et al., 1994). Chloroplastic and mitochondrial targeting signals have some similar features, such as basic pI values (Table II), but the plant TOM complex can still distinguish chloroplastic transit peptides from mitochondrial presequences (Glaser et al., 1998). Compared with chloroplastic stroma transit peptides, our data suggested that formation of an α-helix with low hydrophobicity at the beginning of mitochondrial presequences, presumably caused by a different composition of amino acids such as Arg (R) and Pro (P), could be crucial in distinguishing chloroplastic transit peptides from their mitochondrial counterparts (Table II; Fig. 3C). Our data also supported the previous suggestion that, in general, mitochondrial presequences contained a segment of 12 to 15 amino acids that can form a basic, amphipathic helix (von Heijne, 1986). The basic pI of the presequences from both Arabidopsis and rice mitochondrial proteins could be essential for organelle targeting, potentially via binding with molecular chaperones such as HSP70s (Zhang and Glaser, 2002). Differentiation between mitochondria and chloroplast targeting could be determined by the secondary structure and the relative hydrophobicity at the beginning of the targeting signals, which may be required for binding to TOM or TOC or in the coordination of other molecular chaperones.

In evolutionary terms, the presequences for mitochondrial targeting in yeast could be divergent to those in plants, due to the absence of chloroplasts in the former. This is supported by the observation that yeast mitochondrial proteins had shorter presequences that were less hydrophilic and had lower frequency to form α-helices at the beginning of a presequence (Table II). Interestingly, yeast mitochondrial presequences had a higher frequency of β-sheet formation at the N terminus, in fact as high as that seen in chloroplastic transit peptides (Table II). All of these data are consistent with selectivity in plant mitochondrial target signal evolution that distinguishes them from chloroplastic transit peptides and may have diverged them from the requirements found in yeast.

In plant TOM complexes, the absence of a homolog for yeast Tom22 and changes in its partner receptor, Tom20, revealed a significant rearrangement of the receptor machinery on plant mitochondria (Macasev et al., 2000). Furthermore, there are significant variations in Tom20 between Arabidopsis and rice, the former having four copies in its genome and the latter only one copy, as in yeast (Lister et al., 2007). The diversification of Tom20
sequences in Arabidopsis might allow more variation in mitochondrial presequences than in rice and could be one reason for the longer but less basic and hydrophilic presequences (Table II) and different amino acid compositions (such as R, A, L, and S in Fig. 3) found in Arabidopsis. We could not find any presequence clusters among those identified Arabidopsis and rice proteins using phylogenetic tree analysis (data not shown), consistent with previous claims (Glaser et al., 1998; Sjöling and Glaser, 1998). For Arabidopsis chloroplastic transit peptides, multiple subgroups with distinctive chloroplast-targeting sequence motifs have been revealed recently by site-directed mutagenesis (Lee et al., 2008b). However, these seven groups of distinctive motifs represented less than half of the more than 200 predicted chloroplastic transit peptides (Lee et al., 2008b), indicating that sequence diversity among chloroplastic transit peptides is substantial beyond the groups identified.

One approach to determine the cleavage sites of the plant-specific membrane-bound MPP (Braun et al., 1992; Eriksson et al., 1994; Glaser and Dessi, 1999) is to conduct statistical analysis of plant mitochondrial sequences to define common features around the cleavage sites (Zhang et al., 2001). Our sequence logo analysis revealed that both Arabidopsis and rice mitochondrial cleavage sites could be grouped into three classes, as suggested previously (Zhang et al., 2001), with conserved −3R (class II) and −2R (class I) or without any conserved Arg (class III; Fig. 4). The class II (−3R) group consisted of 55% to 58% of the observed Arabidopsis and rice mitochrondria, indicating that RXK(S/A)(S/T) could be the major motif for MPP processing (Fig. 4). In this class, we additionally show that rice had an even distribution of Y, F, or L in the −1 position, rather than the dominated F observed in Arabidopsis (Fig. 4). The dominance of dicots in the list of presequences used by Zhang et al. (2001) explains the dominant F at the −1 position observed in that study and suggests that species differences in MPP specificity could be very significant among plants. The class I (−2R) group in Arabidopsis appears to have a motif of RX/N/XST, effectively a one-amo acid position shift of class II (−3R; Fig. 4), suggesting no real distinction between the class I and II groups in binding or catalyzing site for MPPs. For the class III group (no conserved R), our data suggested a cleavage motif of (F/Y)/(S/A), which has substantial similarities in the −1 and +1 position to the class III group (−3R; Fig. 4). It is not known whether MPPs or another protease are involved in presequence cleavage of the class III proteins.

Mitochondrial protein import is a complex process with multiple steps, including presequence signal recognition before import, import via TOM and TIM, and presequence cleavage after import. The presequences for mitochondrial protein import in the dicot Arabidopsis and the monocot rice appear to be conserved based on this study, but with some variations, such as in length and amino acid composition of presequences and specific aspects of the cleavage motifs for MPP. Future combinations of MS approaches with analysis by site-directed mutagenesis, in vitro import, and cleavage activity assays could provide deeper insights into the mechanisms of mitochondrial protein import and potentially more detailed pictures of variations between model species. This will aid in a broader understanding of the specialization of the protein targeting to mitochondria required in plants due to the dominant role of plastids and the relative abundance of plastid precursors in plant cells.

MATERIALS AND METHODS

Arabidopsis Culture Cell and Mitochondrial Preparation

A heterothetic Arabidopsis (Arabidopsis thaliana) cell culture was maintained under the same conditions as described by Millar et al. (2001). Arabidopsis cell suspension was cultured in growth medium (1× Murashige and Skoog medium without vitamins, 3% Suc, 0.5 mg L−1 naphthaleneacetic acid, and 0.05 mg L−1 kinetin, pH 5.8) at 22°C under a 16-h-day/8-h-night regime and light intensity of 90 μmol m−2 s−1 with orbital shaking at 120 rpm. Cultures were maintained in 250-mL Erlenmeyer flasks by the inoculation of 25 mL of 7- to 10-d-old cells into 100 mL of fresh growth medium. Dark-grown cells used for mitochondrial isolation were subcultured from 7-d-old light-grown cultures and grown under the same conditions as described for light-grown cell cultures.

Isolation of mitochondria from cell culture was carried out according to the method modified from Millar et al. (2001) and Meyer et al. (2007). Cells (250 g) were incubated at 25°C for 3 h in approximately 800 mL of enzyme buffer (0.4 M mannitol, 36 mM MES, 0.4% [w/v] cellulose [Onozuka; Yakult Pharmaceutical], and 0.05% [w/v] pectolyase [Y-23; Kyowa Chemical Products], pH 5.7). Protoplasts were harvested by washing twice in enzyme buffer without cellulose and pectolyase and centrifugation at 800g for 5 min. Cells were disrupted in ice-cold extraction buffer (0.45 M mannitol, 50 mM tetrasodium pyrophosphate, 0.5% [w/v] polyvinylpyrrolidone [PVP], 0.5% [w/v] bovine serum albumin [BSA], 2 mM EGTA, and 20 mM Cys, pH 8.0) by five strokes in a Potter-Elvehjem homogenizer. The homogenate was then centrifuged at 18,600g for 20 min at 4°C. The pellet of crude organelles was carefully resuspended in ice-cold mannitol washing buffer (0.3 M mannitol, 0.1% [w/v] BSA, and 10 mM TES, pH 7.5). Following one stroke in a Potter-Elvehjem homogenizer, the crude organellar fraction was gently layered over a 35-mL discontinuous Percoll density gradient consisting of 18% (5 mL) over 23% (25 mL) and 40% (5 mL) Percoll solution in mannitol washing medium. The gradient was then centrifuged at 40,000g for 45 min at 4°C with brake off. The mitochondrion fraction was seen as an off-white band near the 23% to 40% (v/v) Percoll interface. The upper layers of the density gradient were removed, and the mitochondrial band was collected. The transferred mitochondrial band was diluted approximately 5-fold with Suc wash medium (0.3 M Suc, 0.1% [w/v] BSA, and 10 mM TES, pH 7.5) and centrifuged at 24,000g for 10 min at 4°C. The mitochondrially enriched homogenate was collected, and Percoll density gradient centrifugation was repeated once as described for the first gradient. Afterward, the mitochondrial sample was washed three times at 21,000g for 20 min each at 4°C using the Suc wash medium as described above but not containing BSA. The amount of mitochondrial protein was determined (Bradford, 1976).

Growth of Rice Seedlings and Mitochondrial Preparation

Batches of 200 g of rice (Oryza sativa ‘Amaroo’) seeds were washed in 1% (w/v) bleach for 10 min, rinsed in distilled water, grown in the dark in vermiculite trays (30 × 40 cm) at a constant 30°C, and watered daily, and the shoots were harvested at 10 d for mitochondrial isolation (Heazlewood et al., 2003). Rice shoots were harvested using scissors and cut into 5- to 10-mm lengths into a beaker on ice. Cut shoots (100 g) were ground with a precooled mortar and pestle using acid-washed sand in 300 mL of homogenization solution (0.3 M Suc, 25 mM tetrasodium pyrophosphate, 2 mM EDTA, 10 mM
KH$_2$PO$_4$, 1% [w/v] PVP-40, 1% [w/v] BSA, and 20 mM ascorbate, pH 7.5). After filtering the homogenate through four layers of Miracloth, it was centrifuged for 5 min at 1,000g at 4°C, and the supernatant was centrifuged again at 20,000g for 20 min at 4°C. The resultant organelle pellet was resuspended in wash buffer (0.3 M Suc, 10 mM TES-KOH, pH 7.5, and 0.1% [w/v] BSA), and the two centrifugation steps were repeated to produce a washed organelle pellet. This resuspended pellet was layered over a 0% to 4.4% (v/v) PVP-40 preformed gradient in a 28% (v/v) Percoll self-forming gradient in wash buffer and centrifuged at 30,000g for 45 min at 4°C with brake off. Mitochondria formed a band toward the bottom of the gradient, the upper plastid material was discarded, and the mitochondrial band was removed in an attempt to minimize contamination of peroxisomal material from the bottom of the gradient. After 0% to 4.4% Percoll gradient centrifugation, mitochondrial samples were washed three times with washing medium (0.3 M Suc and 10 mM TES, pH 7.5) and centrifuged at 21,000g for 20 min at 4°C. The amount of mitochondrial protein was determined according to the method of Bradford (1976).

Two-Dimensional Gel Electrophoresis
Mitochondrial protein samples from Arabidopsis and rice (700 µg) were extracted by addition of cold acetone (~20°C) to a final concentration of 80% (v/v). Samples were stored at ~80°C for 4 h and then centrifuged at 20,000g at 4°C for 15 min. The pellets were resuspended in IEF sample buffer (7 M urea, 2% TCA, 4% [w/v] CHAPS, and 40 mM Tris base, pH 8.5). Aliquots of 450 µL were used to reswell immobilized pH gradient strips (pH3-10 NL; 24 cm; GE Healthcare) according to the manufacturer’s instructions. The strips were run for 24 h in Ettan IPGphor3 (GE Healthcare) according to the manufacturer’s instructions. The strips were then transferred to an equilibration buffer (50 mM Tris-HCl [pH 6.8], 4 M urea, 2% [w/v] SDS, 0.001% [w/v] bromphenol blue, and 100 mM β-mercaptoethanol) and incubated for 20 min at room temperature with rocking. After a brief wash in 1X gel buffer, the strips were transferred on top of 12% acrylamide Gly gels and covered with 1.2% (v/w) agarose in gel buffer. Second-dimension gels were run at 50 mA per gel for 6 h. Proteins were visualized by colloidal Coomassie Brilliant Blue G250 staining.

Analysis of Peptides from In-Gel-Digested Protein Samples
After Arabidopsis and rice mitochondrial proteins were separated by two-dimensional IEF-SDS-PAGE, protein samples to be analyzed were cut from the gels and in-gel digested with trypsin according to the method described by Taylor et al. (2005). Samples were then dried in a vacuum centrifuge, resuspended in 5% (v/v) acetonitrile and 0.1% (v/v) formic acid, and analyzed on an Agilent XCT Ultra Ion Trap mass spectrometer (Agilent Technologies) according to Meyer et al. (2007). Resulting MS/MS-derived spectra were analyzed against our in-house rice mitochondrial protein database mainly extracted from our recently generated rice mitochondrial protein database (Huang et al. 2009) and our in-house Arabidopsis mitochondrial protein database extracted from SUBA (Heazlewood et al., 2007; http://www.plantenergy.uwa.edu.au/applications/suba/index.php). The database was searched using the Mascot search engine version 2.2.03 with nonenzyme digestion and mass error tolerances of ±100 ppm for MS and ±0.5 D for MS/MS. Mass Spectral evidence for N-terminal acetylation (N-term), Carbamidomethyl (C), and Acetylation (N-term), instrument set to ESI-QUAD-TOF, and peptide charge set at 2+ and 3+. Results were filtered using “mudpit scoring,” “Max. number of hits” set to 400, and “ion score cut-off” at 20. The significance threshold P ≤ 0.05 and “Require bold red” were also set.

Analysis of Peptides from Whole Mitochondrial Digests
Whole Arabidopsis and rice mitochondrial protein extracts were digested overnight at 37°C in the presence of trypsin, and insoluble components were removed by centrifugation at 20,000g for 5 min. Samples were analyzed on an Agilent 6510 Q-TOF mass spectrometer with an HPLC Chip Cube source (Agilent Technologies). The chip consisted of a 40-nL enrichment column (Zorbax 300SB-C18; 5 µm pore size) and a 150-mm separation column (Zorbax 300SB-C18; 5 µm pore size) driven by the Agilent Technologies 1100 series nano/capillary liquid chromatography system. Both systems were controlled by MassHunter Workstation Data Acquisition for Q-TOF (version B.01.02, Build 65.4, Patch 1.2.3.4; Agilent Technologies). Peptides were loaded onto the trapping column at 4 µL min$^{-1}$ in 5% (v/v) acetonitrile and 0.1% (v/v) formic acid with the chip switched to enrichment and using the capillary pump. The chip was then switched to separation, and peptides eluted during a 1-h gradient (5% [v/v] acetonitrile to 40% [v/v] acetonitrile) directly into the mass spectrometer. The mass spectrometer was run in positive ion mode, and MS scans were run over a mass-to-charge ratio range of 275 to 1,500 and at four spectra s$^{-1}$. Precursor ions were selected for auto MS/MS at an absolute threshold of 500 and a relative threshold of 0.01, with a maximum of three precursors per cycle, and active exclusion set at two spectra and released after 1 min. Precursor charge-state selection and preference was set to 2+ and then 3+. Results were selected by charge and then abundance. Resulting MS/MS spectra were opened in MassHunter Workstation Qualitative Analytical Evaluation (version B.01.02, Build 1.2.122.1, Patch 3; Agilent Technologies), and MS/MS compounds were detected by “Find Auto MS/MS” using default settings. The resulting compounds were then exported as mzdata files, which when appropriate were combined using mzdata Combinator version 1.0.4 (West Australian Centre of Excellence in Computational Systems Biology; http://www.plantenergy.uwa.edu.au/waccecb/software.shtml). Searches were conducted using Mascot Search Engine version 2.2.03 (Matrix Science) with nonenzyme digestion and mass error tolerances of ±100 ppm for MS and ±0.5 D for MS/MS. Mass Spectral evidence for N-terminal acetylation (N-term), Carbamidomethyl (C), and Acetylation (N-term), instrument set to ESI-QUAD-TOF, and peptide charge set at 2+ and 3+. Results were filtered using “mudpit scoring,” “Max. number of hits” set to 400, and “ion score cut-off” at 20. The significance threshold P ≤ 0.05 and “Require bold red” were also set.

Bioinformatics Tools
A prediction of the cleavage sites of mitochondrial proteins was performed using the Web server of TargetP (http://www.cbs.dtu.dk/services/TargetP/). The pl values of mitochondrial presequences and chloroplast stromal transit peptide were analyzed using the Compute pl/Mw tool (http://au.expasy.org/tools/pltool.html). The hydrophobicity values of amino acid residues in protein were analyzed using the Web server TopPred program (http://moby.proteum.fr/cgi-bin/MobyPortal/portal.py?form=toppred; Clarois and von Heijne, 1994). The averaged hydrophobicity values of the first 10 amino acids in proteins are presented in Supplemental Tables S2 to S4. The secondary structures of the obtained mitochondrial presequences and chloroplastic stromal transit peptides were determined using the Web server JPred3 (http://www.compbio.dundee.ac.uk/~www-jpred;/ Cole et al., 2008). H and E denoted the secondary structures of α-helix and β-sheet, respectively. Sequence logos were created using the Web server Weblogo (http://weblogo.berkeley.edu/logo.cgi; Crooks et al., 2004). Fifteen C-terminal residues of the presequences and the 15 residues from the N terminus of the mature proteins were used for sequence logo analysis.

Statistical Analysis
The significance of differences between proportions was analyzed using RatioDiff, which computes P values based on the comparison of proportions in paired samples (http://www.cecks.org/applications/ratiodiff/).

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

Supplemental Figure S1. Mass spectral evidence for N-terminal acetylation of Arabidopsis mitochondrial GDH.

Supplemental Figure S2. Frequency distribution of the predicted secondary structures of the 15 amino acid residues before and after cleavage sites of the observed 62 Arabidopsis and 52 rice mitochondrial proteins.

Supplemental Table S1. Comparison of the MS-identified semimetric peptides with reported N-terminal sequences deduced from Edman degradation (Kraft et al., 2001).

Supplemental Table S2. The 62 Arabidopsis and 52 rice mitochondrial presequences derived from MS evidence.
**LITERATURE CITED**


Braun HP, Emmermann M, Kruft V, Schmitz UK (1992) The general mitochondrial processing peptidase from potato is an integral part of cytochrome c reductase of the respiratory chain. EMBO J 11: 3219–3227


Julliard JH, Smith EL (1979) Partial amino acid sequence of the glutamate dehydrogenase of human liver and a revision of the sequence of the bovine enzyme. J Biol Chem 254: 3427–3438


Meyer EH, Heazlewood JH, Millar AH (2007) Mitochondrial acyl carrier proteins in Arabidopsis thaliana are predominantly soluble matrix proteins and none can be assigned as subunits of respiratory complex I. Plant Mol Biol 64: 319–327


**Supplemental Table S3.** The 105 yeast mitochondrial proteins with experimentally proven presequences listed in the Swiss-Prot database.

**Supplemental Table S4.** The 47 Arabidopsis stroma presequences derived from MS evidence of N-terminal acetylated semitryptic peptides from Zybailov et al. (2008).

**Supplemental Table S5.** The 57 experimentally proven plant mitochondrial presequences from the Swiss-Prot database (A) and other experimentally proven plant mitochondrial presequences that are not classified in the Swiss-Prot database (B).

**Supplemental Table S6.** The 19 ortholog sets of presequences with rice and Arabidopsis members derived from data in Supplemental Table S2.
degrading targeting peptides in mitochondria and chloroplasts. Plant J 36: 616–628
Roise D, Horvath SJ, Tomich JM, Richards JH, Schatz G (1986) A chemically synthesized pre-sequence of an imported mitochondrial protein can form an amphipilic helix and perturb natural and artificial phospholipid bilayers. EMBO J 5: 1327–1334