SUPPLEMENTAL APPENDIX

Most plant cytosolic mRNAs have a 5'-terminal cap comprising a 7-methylguanosine residue joined by a 5'-to-5' triphosphate linkage. A cap-binding complex recruits the 40S ribosomal subunit to the 5' end of the message, which then scans along the transcript until the first AUG codon is encountered, at which point the 60S ribosomal subunit also binds and translation commences (Kozak, 1989). While this "ribosome scanning" mechanism of translation initiation is used with most transcripts, exceptional modes of cap-dependent and cap-independent initiation do operate, especially in conjunction with viral mRNAs (Touriol et al., 2003). One mechanism allowing for the use of an alternative, downstream AUG codon is "leaky scanning". In this case, the ribosome passes over the first AUG, because it is in an unfavourable context, and instead utilizes the next, more favourable AUG codon for translation initiation. In the case of *AtOEP80*, the first AUG codon is actually more similar (8/11 matches) with the dicot consensus (5'-aaA(A/C)aAUGGCu-3'; Joshi et al., 1997) than the second AUG (6/11 matches).

Secondary structural elements in transcript leader sequences can also influence the position of initiation (Chabregas et al., 2003; Touriol et al., 2003). To assess whether such factors might play a role in AtOEP80 translation, we examined the 5' end of the transcript in silico (using DNASTAR GeneQuest software, version 7.2). A stem-loop structure was predicted to occur adjacent to AUG1, while no such structures were predicted in the vicinity of AUG2. A more thorough analysis using specialized software (mfold version 3.2; Mathews et al., 1999; Zuker, 2003) supported the initial AUG1 stem-loop prediction, and revealed that it may be sufficiently stable to interfere with translation (Supplemental Fig. S4A). If the AUG1 stem-loop does indeed impede canonical initiation, there should be a mechanism for internal ribosome initiation further downstream. Previous reports indicated that complementarity between leader sequences and the central region of the 18S rRNA can act to recruit ribosomes in cap-independent fashion (Akbergenov et al., 2004; Dresios et al., 2006). We examined the *AtOEP80* leader for such an internal ribosome entry site (IRES), and found a 17 nucleotide stretch, just upstream of AUG2, that is complementary to the Arabidopsis 18S rRNA (Supplemental Fig. S4B). Interestingly, this region of complementarity overlaps well with a sequence ("1105-1114"; Supplemental Fig. S4C) shown previously to strongly promote cap-independent binding and enhance translational efficiency (Akbergenov et al., 2004).

Use of a non-canonical, downstream initiation might have developmental or regulatory significance, enabling functional differentiation or optimization of the protein. It is interesting to note that in *Neurospora crassa*, a mitochondrial Omp85 homologue, Tob55 (Topogenesis of β-barrel proteins, 55 kD; an alternative name for Sam50 [Sorting and assembly machinery, 50 kD]), exists in three different forms due to alternative splicing, and that these exhibit functional differences (Hoppins et al., 2007). Another possibility is that downstream initiation might optimize the organellar targeting properties of AtOEP80 in vivo. Analysis of the long (732 residues) and short (680 residues) forms of

the protein using some localization prediction programs supported this hypothesis (Schein et al., 2001; Bannai et al., 2002; Small et al., 2004; Horton et al., 2007), although in vitro import data are inconclusive in this regard (Fig. 5A).

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure S1. Expression patterns of the *AtOEP80* gene relative to genes for other envelope proteins.

Publicly available Affymetrix GeneChip microarray data were analyzed and retrieved using the Genevestigator V3 analysis tool (https://www.genevestigator.ethz.ch/) (Zimmermann et al., 2004; Grennan, 2006). Presented data were prepared using the Meta-Profile Analysis tool, using either the Development (A) or Anatomy (B) representations in scatter-plot format. Data from all high-quality, ATH1(22k) arrays were analyzed; this amounted to a total of 3110 arrays. Values shown are means (± SE). The total number of arrays used to derive each data point shown is indicated in each panel. The software differentiates between signals that are significantly above background noise, and those which are close to background: probe sets represented by closed circles are those which are called "present", while those represented by open circles (none in this case) are called "absent". Data representations were exported from Genevestigator in Encapsulated PostScript format, and then compiled and annotated using appropriate graphics software. The genes analyzed were as follows: *atTOC75-III* (At3g46740; red); *atTIC110* (At1g06950; blue); *AtOEP80* (At5g19620; yellow); *atTOC75-IV* (At4g09080; green).

Supplemental Figure S2. Annotated T-DNA left border junction sequence for the *oep80-3* mutant. Sequence flanking the LB junction of the *oep80-3* T-DNA insertion was amplified as illustrated in Figure 2B. The composition of the amplified fragment was analysed using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) provided by the National Center for Biotechnology Information (NCBI), and then annotated accordingly. Orange bars show the location of DNA sequences corresponding to the LB region of the T-DNA, and to the 5' region of the *AtOEP80* gene. The green bar shows the conceptual position of the *AtOEP80* transcript synthesized in the *oep80-3* mutant, as determined by RACE-PCR; the blue bar shows the position of the open reading frame within the transcript, and the red bar shows the sequence of the putative AtOEP80 protein, initiating from the second AUG codon of the wild-type *AtOEP80* transcript (nucleotides 276 to 278 of cDNA NM_121967).

Supplemental Figure S3. Specificity and affinity purification of the AtOEP80 antibody.
A, Confirmation of the specificity of the AtOEP80 antiserum by competition assay. Protein samples from chloroplasts (equivalent to 10 μg chlorophyll) of either wild-type ("WT") or homozygous *oep80-3* mutant ("*80-3*") Arabidopsis seedlings, were separated by SDS-PAGE, transferred to PVDF membrane, and incubated with crude antiserum against AtOEP80 (Inoue and Potter, 2004; left panel), or the same antiserum pre-incubated with 10 μg/mL of the antigen peptide (right panel). Protein bands corresponding to AtOEP80 are visible in the left panel, as indicated, but not in the right panel. Positions of molecular weight standards are indicated at left (sizes in kD).
B, Affinity purification of the AtOEP80 antibody. The crude anti-peptide antiserum described in A above, and previously (Inoue and Potter, 2004), was purified using UltraLink Iodoacetyl Gel coupled with the antigen peptide. Arabidopsis chloroplast protein samples (equivalent to 10 μg chlorophyll) were separated by SDS-PAGE, and then analyzed by immunoblotting using the crude serum, as well as unbound and eluted fractions from the purification column. The position of the bands corresponding to AtOEP80 are indicated at right. Positions of molecular weight standards are indicated at left (sizes in kD).

Supplemental Figure S4. In silico analysis of the 5' region of the *AtOEP80* mRNA.
A, Stable stem-loop structure at the first AUG codon. Sequence spanning the first AUG codon of the native *AtOEP80* transcript (nucleotides 118 to 144 of cDNA NM_121967) was analyzed using the mfold version 3.2 web server (http://www.bioinfo.rpi.edu/applications/mfold/) (Mathews et al., 1999; Zuker, 2003). Hybridization of bases is indicated as follows: red, CG or GC base pairing; blue, AU or UA base pairing; green, GU or UG wobble base pairing. Nucleotides are annotated with the results of ss-count analysis (see key): this measures the propensity of a base to be single stranded, as determined by the number of times it is single stranded in a group of predicted foldings. The calculated minimum free energy associated with the folded structure, ΔG, is indicated.
B, Complementarity between the 5' end of the *AtOEP80* transcript and 18S rRNA. Sequence between AUG1 and AUG2 of *AtOEP80* (nucleotides 120 to 278 of cDNA NM_121967) and nucleotides 1001 to 1200 of Arabidopsis 18S rRNA (At3G41768) were analysed together using mfold. This identified the region of complementarity shown, which was then re-analysed using shorter sequences, together with an appropriate linker (LLL), in order to calculate the free energy associated with the hybridized structure. Hybridization of bases is indicated as in A.
C, Overlap of the 18S-rRNA-complementary region with sequence motifs known to promote cap-independent binding and enhance translation efficiency. The rRNA-complementary region of *AtOEP80* shown in B is aligned with reverse complements (RC) of 18S rRNA sequences from Arabidopsis (At18S) and rice (Os18S); numbers at right refer to positions in the forward complement in each case. Complementarity between *AtOEP80* and the 18S sequences is indicated by shading,

using the same colour scheme as in A and B: red, CG or GC base pairing; blue, AU or UA base pairing; green, GU or UG wobble base pairing. The positions of translation enhancer motifs identified by Akbergenov et al. (2004) are indicated below the alignment.

SUPPLEMENTAL TABLES

Supplemental Table SI. *Segregation of the T-DNA-associated selectable marker in each of the oep80 mutants*

Supplemental Table SII. *Phenotypic analysis of the oep80 mutants*

SUPPLEMENTAL LITERATURE CITED

Akbergenov R, Zhanybekova S, Kryldakov RV, Zhigailov A, Polimbetova NS, Hohn T, Iskakov BK (2004) ARC-1, a sequence element complementary to an internal 18S rRNA segment, enhances translation efficiency in plants when present in the leader or intercistronic region of mRNAs. Nucleic Acids Res. **32:** 239-247

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. **215:** 403-410

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18:** 298-305

Chabregas SM, Luche DD, Van Sluys MA, Menck CF, Silva-Filho MC (2003) Differential usage of two in-frame translational start codons regulates subcellular localization of *Arabidopsis thaliana* THI1. J. Cell Sci. **116:** 285-291

Dresios J, Chappell SA, Zhou W, Mauro VP (2006) An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. Nat. Struct. Mol. Biol. **13:** 30-34

Grennan AK (2006) Genevestigator. Facilitating web-based gene-expression analysis. Plant Physiol. **141:** 1164-1166

Hoppins SC, Go NE, Klein A, Schmitt S, Neupert W, Rapaport D, Nargang FE (2007) Alternative splicing gives rise to different isoforms of the *Neurospora crassa* Tob55 protein that vary in their ability to insert beta-barrel proteins into the outer mitochondrial membrane. Genetics **177:** 137-149

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res. **35:** W585-W587

Inoue K, Potter D (2004) The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. Plant J. **39:** 354-365

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol. Biol. **35:** 993-1001

Kozak M (1989) The scanning model for translation: an update. J. Cell Biol. **108:** 229-241

Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288:** 911-940

Schein AI, Kissinger JC, Ungar LH (2001) Chloroplast transit peptide prediction: a peek inside the black box. Nucleic Acids Res. **29:** E82

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for *N*-terminal targeting sequences. Proteomics **4:** 1581-1590

Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol. Cell **95:** 169-178

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol. **136:** 2621-2632

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:** 3406-3415