https://doi.org/10.1093/plphys/kiad572 PLANT PHYSIOLOGY 2024: 194: 1411–1430

Plant Physiology®

Research Article

Detection and editing of the updated Arabidopsis plastid- and mitochondrial-encoded proteomes through PeptideAtlas

Klaas J. van Wijk [1,*] Stephane Bentolila [2,*] Tami Leppert [3] Qi Sun [4] Zhi Sun [3] Luis Mendoza [3] Margaret Li [3] and Eric W. Deutsch [3,*]

1 Section of Plant Biology, School of Integrative Plant Sciences (SIPS), Cornell University, Ithaca, NY 14853, USA
2 Department of Molecular Biology & Genetics, Cornell University, Ithaca, NY 14853, USA
3 Institute for Systems Biology (ISB), Seattle, WA 98109, USA
4 Computational Biology Service Unit, Cornell University, Ithaca, NY 14853, USA

*Author for correspondence: kv35@cornell.edu (K.J.v.W.), sb46@cornell.edu (S.B.), edeutsch@systemsbiology.org (E.W.D.)
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Klaas J. van Wijk (kv35@cornell.edu).

Abstract

Arabidopsis (*Arabidopsis thaliana*) ecotype Col-0 has plastid and mitochondrial genomes encoding over 100 proteins. Public databases (e.g. Araport11) have redundancy and discrepancies in gene identifiers for these organelle-encoded proteins. RNA editing results in changes to specific amino acid residues or creation of start and stop codons for many of these proteins, but the impact of RNA editing at the protein level is largely unexplored due to the complexities of detection. Here, we assembled the nonredundant set of identifiers, their correct protein sequences, and 452 predicted nonsynonymous editing sites of which 56 are edited at lower frequency. We then determined accumulation of edited and/or unedited proteoforms by searching ~259 million raw tandem MS spectra from ProteomeXchange, which is part of PeptideAtlas (www.peptideatlas.org/builds/arabidopsis/). We identified all mitochondrial proteins and all except 3 plastid-encoded proteins (NdhG/Ndh6, PsbM, and Rps16), but no proteins predicted from the 4 ORFs were identified. We suggest that Rps16 and 3 of the ORFs are pseudogenes. Detection frequencies for each edit site and type of edit (e.g. S to L/F) were determined at the protein level, cross-referenced against the metadata (e.g. tissue), and evaluated for technical detection challenges. We detected 167 predicted edit sites at the proteome level. Minor frequency sites were edited at low frequency at the protein level except for cytochrome C biogenesis 382 at residue 124 (Ccb382-124). Major frequency sites (>50% editing of RNA) only accumulated in edited form (>98% to 100% edited) at the protein level, with the exception of Rpl5-22. We conclude that RNA editing for major editing sites is required for stable protein accumulation.

Introduction

Arabidopsis (*Arabidopsis thaliana*) and all other plants have plastid and mitochondrial genomes encoding proteins, several tRNAs, and rRNAs and also include 4 tentative ORFs. These organelle-encoded proteins are components of the protein complexes of the plastid and mitochondrial electron transport chains, components of the transcriptional and translations machineries, protein biogenesis factors, and several proteins involved in metabolism (Green 2011; Zoschke and Bock 2018; Møller et al. 2021). The precise function of just a few organelle-encoded proteins is still unknown (e.g. plastid hypothetical chloroplast ORF 1 and 2 [YCF]), whereas the 4 ORFs might be pseudogenes (i.e. plastid-encoded Ycf15 and mitochondrial OrfX/TatC/MttB, Orf114, and Orf240A). Most sources report around 88 protein-coding plastid genes and 31 to 33 protein-coding

Received July 10, 2023. Accepted September 23, 2023. Advance access publication October 25, 2023

mitochondrial genes in Arabidopsis, based on initial sequenced genomes for plastids for ecotype Columbia (Col-0; Sato et al. 1999) and mitochondria for ecotype C24 (Unseld et al. 1997). The Araport11 genome assembly (and also the prior assembly TAIR10) includes 122 mitochondrial-encoded protein identifiers (ATMG) rather than the expected 31 to 33 (Sloan et al. 2018). Furthermore, several plastid (ATCG)- or mitochondrial (ATMG)-encoded protein identifiers in Araport11/TAIR10 represent individual exons that are trans-spliced to make complete protein-coding transcripts (i.e. plastid-encoded Rps12 and mitochondrial-encoded Ndh1,2,5) or are redundant because they represent duplicated copies of protein-coding genes located on the inverted repeats (IRs) of the plastid genome. Additionally, many of the organelle-encoded proteins have more than 1 protein name in the literature and various databases. The current study aims to first clarify this confusion and provides a consensus nonredundant set of plastid- and mitochondrial-encoded proteins with their amino acid sequences for ecotype Col-0, including the protein identifiers (ATCG and ATMG). We suggest that this updated set be incorporated in the forthcoming Arabidopsis genome annotation for Col-0 (tinyurl.com/Athalianav12).

Several plastid- and most mitochondrial-encoded mRNAs undergo C to U mRNA editing, which can affect the resulting protein sequence, i.e. nonsynonymous edits (Takenaka et al. 2013; Germain et al. 2015; Fuchs et al. 2020; Small et al. 2020). In the case of NdhD/Ndh4, mRNA editing results in introduction of a start codon (thus generating a translatable mRNA), and in 10 cases (within Nad9, Rpl16, Rps3, Rps12, CsmB, CcmF, MatR, and TatC), mitochondrial gene editing introduces an extra stop codon, which would result in a truncated protein. It is generally believed that RNA editing events are required for protein function and/or protein stability (Small et al. 2023). The extent of quantitative editing for each site has been reported at the RNA level based on RNA-seq data, and several editing sites showed variability in the extent of editing (partial editing) depending on growth or stress conditions, developmental state, and/or tissue type (Ruwe et al. 2013; Germain et al. 2015). The impact of RNA editing on proteins is best determined by proteomics and tandem MS (MS/MS). However, very little systematic information about editing is available at the protein level, mostly because of technical challenges and low sequence coverage. In a typical proteome experiment, the isolated proteome is converted into peptides using enzymatic digestion with a protease (typically trypsin), followed by MS/MS analysis of the peptide mixture. The technical challenges to map editing sites at the protein level include incomplete protein sequence coverage, in particular for regions that are very hydrophobic (transmembrane domains) or with many basic amino acids (Lys, Arg). An added challenge is accommodation of the protein search space to allow for detection of closely spaced editing sites in the same protein independently of each other.

In this study, we determined the accumulation of unedited and edited isoforms of the plastid- and mitochondrial-encoded proteins and possible partial editing by searching ~259 million MS/MS spectra as part of the Arabidopsis PeptideAtlas project (van Wijk et al. 2021; van Wijk et al. 2023). To facilitate the protein search and optimize the protein search space, we reached out to members of the plant community and gathered literature annotations (e.g. Germain et al. 2015; Lenz et al. 2018) to obtain the most complete set of possible organelle-encoded proteins and their amino acid sequences, including their unedited and edited variants. Whereas we identified most organelle-encoded proteins, no MS/MS support was found for proteins encoded by the ORFs. We report on the detection of edited sites and make recommendations for protein search space and search strategies for future proteomics studies that include the plastid- and mitochondrial-encoded proteomes. Finally, we also report on 4 important physiological posttranslation modifications of the organelle-encoded proteins in PeptideAtlas build 2, namely N-terminal acetylation (NTA), lysine acetylation, and phosphorylation, as well as ubiquitination, based on raw MS data from submissions in ProteomeXchange that specifically enriched for these PTMs.

Results

Assembly of the nonredundant plastid- and mitochondrial-encoded proteins, protein sequences, and predicted editing sites

To arrive at a consensus set of plastid- and mitochondrial-encoded protein sequences and their predicted nonsynonymous RNA editing sites, we collected information from a wide range of public resources (see Materials and methods). Furthermore, we reached out to the plastid and mitochondrial research community to solicit input on plastid and mitochondrial editing sites as well as physiological relevance of predicted organellar ORFs (see Acknowledgments).

Supplemental Data Set 1 summarizes the plastid-encoded proteins with their gene ATCG identifier, protein name and function, and nonsynonymous editing events. These identifiers and their unedited and (partially) edited amino acid sequence variants can be downloaded at Arabidopsis PeptideAtlas. There are 88 ATCG identifiers in TAIR (identical across Araport11 and TAIR10) due to various types of redundances (see below), but there are only 79 unique full-length protein sequences. Within the most recent release of the Arabidopsis PeptideAtlas and in this study, we removed this redundancy (Supplemental Data Set 1). For 4 proteins (MatK, NdhB/Ndh2, PsbC, and Rps19), there are conflicting database entries about the start of the protein sequence. In the case of MatK, the shorter form (22 amino acids shorter) appears correct (N-terminus is MDKFQGYLEF) and not MCHFRTQENKDFTFSSNRISIQ as also supported by lack of observed matched peptides to the extended N-terminus (Fig. 1A). In the case of NdhB/Ndh2, it is not clear if the longer form (N-terminus is MIWHVQNENF; 123 amino acids longer) is more accurate that the N-terminally shorter form (N-terminus MAITEFLLF; Supplemental Data Set 1). In

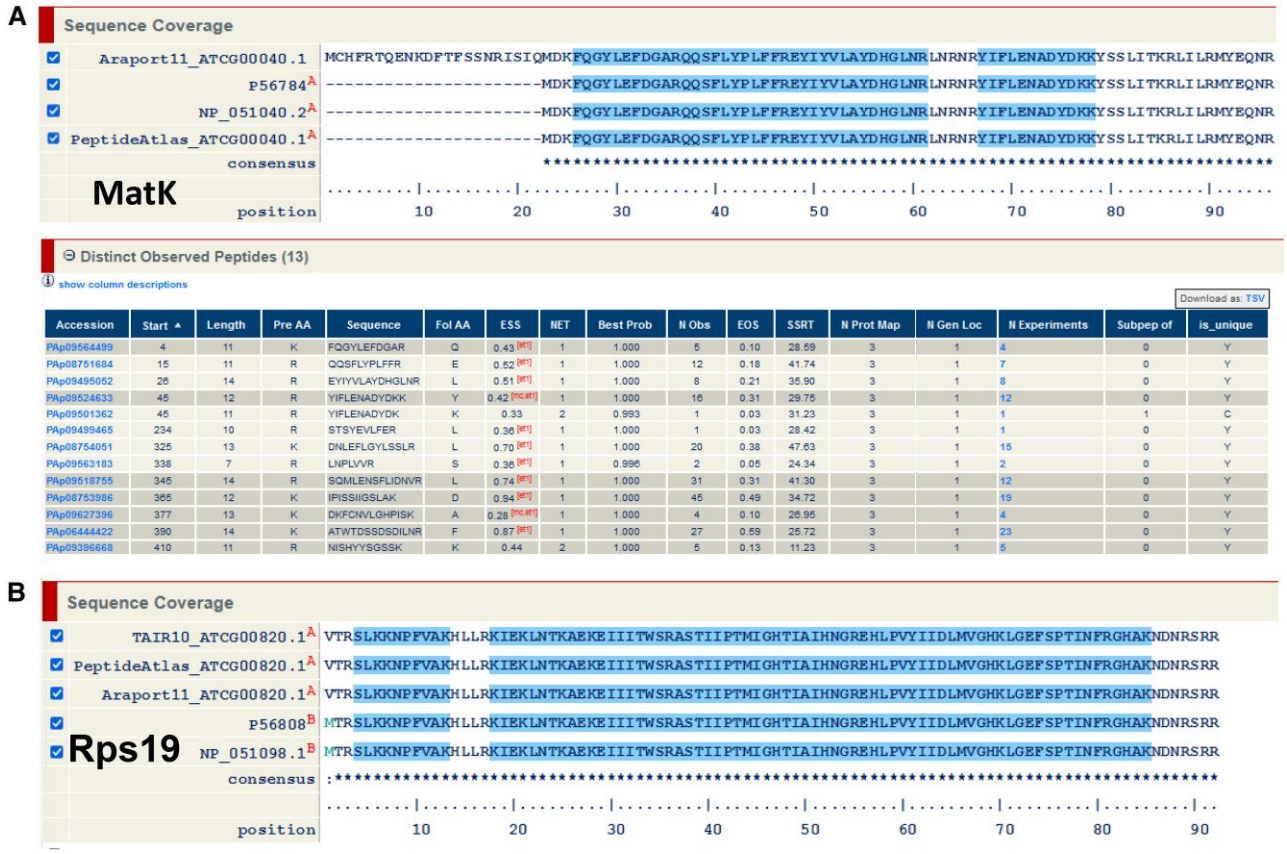


Figure 1. Evaluation of conflicting N-terminal annotation of MatK and Rps19. **A)** Primary protein sequence for MatK and conflicting reports on its N-terminus. The shorter form (22 amino acids shorter) appears correct (N-terminus is MDKFQGYLEF) and not MCHFRTQENKDFTFSSNRISIQ as also supported by lack of observed matched peptides to the extended N-terminus. **B)** Primary protein sequence for Rps19 and conflicting reports on its N-terminus. Rps19 starts with VTRSLKKNPFVAKHLL in the case of ATCG00820 (in PeptideAtlas, TAIR10, and Araport11), but the N-terminal residue is methionine in the case of UniProt P56808 and NP_051098. PeptideAtlas has matching peptides for tryptic peptide SLKKNPFVAK, which does not provide clarification on the N-terminal residue. However, initiation with formylmethionine on GUG codons via wobble pairing with tRNA-fMet(CAU) is well known in bacteria and long suspected in organelles.

fact, a third N-terminus (MKAFHLLL) is possibly located in between these 2 N-termini and was generated by RNA editing in other plant species (Supplemental Fig. S1). In the case of Rps19, the N-terminus is VTRSLK in the case of ATCG00820 (in PeptideAtlas, TAIR10, and Araport11) but the N-terminal residue is methionine (MTRSLK) in the case of UniProt P56808 and NP_051098 (Fig. 1B). PeptideAtlas has matching peptides for tryptic peptide SLKKNPFVAK, which does not provide clarification on the N-terminal residue (Fig. 1B), and the true N-terminus of Rps19 is unclear. However, initiation with formylmethionine on GUG codons via wobble pairing with tRNA-fMet(CAU) is well known in bacteria and long suspected in organelles. Indeed, a recent paper with stable transplastomic tobacco (*Nicotiana tabacum*) plants showed that GUG can indeed be used as a start codon in chloroplasts (Sadhu et al. 2023). Therefore, Rps19 likely starts with MTRSLL as suggested in UniProt P56808 and NP_051098. PsbC has been shown to start at a GUG codon in tobacco (Kuroda et al. 2007). With similar reasoning as for Rps19, this could result in a N-terminal methionine. Indeed, all protein models in PeptideAtlas, Araport, and UniProt show the same N-terminal sequence MKTLYSLRRFYHVETLFNGTLALAGR for PsbC (Supplemental Fig. S2). There are 277 different peptides in PeptideAtlas with a total of 262,161 peptide-spectrum matches (PSMs) that mapped to PsbC. Hence, there is a vast amount of peptide information, but peptides that are observed with very low frequency are possibly false discoveries. The most N-terminal peptides in PeptideAtlas for PsbC (Peptide Atlas_ATCG00280) are $\mathbf{R_9}$FYHVETLFNGTLALAGRDQETT GFAWWAGNAR (2 missed cleavages) with only 2 PSMs (out of >262,161) and tryptic peptide F_{10}YHVETLFN GTLALAGR with 11 observations. Hence, there is no coverage for the first 6 residues (MKTLYSLR)—possibly because the full tryptic peptide (TLYSLR) is too short (less than 7 aa residues) to be considered in the MS/MS search. After

methionine excision and tryptic digestion, the peptide would be TLYSLR, which is only 6 residues. In the case of a missed cleavage, the peptide would be KTLYSLR (7 aa), which is just long enough for detection. There are peptides matching to PsbC starting at every position from residues 9 to 20. However, the frequency is very low (given a total number of PSMs of 262,161) for positions 9, 11, 12, and 13 (between 1 and 4), 11 for position 10, 12 for position 14 but very high at 493 for position 15 (TLFNGTLALAGR—semitryptic). Inspection of the matched spectra for these N-terminal peptides starting at positions 8 to 15 indicates that spectral matches for peptides starting at positions 8 and 9 are very atypical, with complete y-ion series for the C-terminal portion but very few b-ions matching to the N-terminal portion of the peptides. Similarly, there are complete y-ion series for the C-terminal residues in FYHVETLFNGTLALAGR but very few matched ions for the N-terminal portion. This suggests these PSMs are false matches. In contrast, near-complete y- and b-ion series were observed for peptides starting at position 14 (E), some of which were phosphorylated at position T15. Inspection of internal data in our Plant Proteome Database (PPDB at http://ppdb.tc.cornell.edu/) shows that most of the N-terminal peptide started at position 15 (TLFNGTLALAGR) and was observed with thousands of PSMs; more than 2,000 PSMs showed acetylation of this N-terminal threonine. In conclusion, by far the most frequently observed N-terminus starts at residue T15 (downstream of E)—given that this is a semitryptic peptide and that it is very often acetylated, this most likely represents the mature N-terminus of PsbC. It is however possible that this N-terminus was generated by a cleavage event and that the synthesized PsbC protein starts more upstream.

For Rps12, there are 3 ATCG identifiers each representing an exon—and we selected the identifier with the lowest number (ATCG00065) and associated the full-length protein sequence with this identifier. Six pairs of ATCG identifiers describe identical duplicates of proteins expressed from genes in the 2 IRs (Rpl2, Rpl23, Rps7, Ycf15, Ycf2, and NdhB). For each pair, we selected the identifier with the lowest number and excluded the others. Finally, ATCG01000 represents a truncated, nonfunctional Ycf1 ORF, and this entry can also be removed for proteome analysis (and there are no peptides that map uniquely to the ATCG01000 sequence). In the case of 17 (unique) proteins, mRNA editing can result in 1 or more amino acid changes. In total, predicted editing would result in 31 amino acid changes and generation of 1 start methionine in the case of NdhD/Ndh4 (Supplemental Data Set 1).

Supplemental Data Set 2 summarizes the mitochondrial-encoded proteins with their ATMG gene identifier and protein name and editing events that impact the protein sequence. These identifiers and their unedited and (partially) edited amino acid sequence variants can be downloaded at Arabidopsis PeptideAtlas. This table is mostly based on mitochondrial genome sequences for Arabidopsis ecotypes C24 and Col-0 (Unseld et al. 1997; Davila et al. 2011), the extensive body of mitochondrial literature (e.g. Rao et al. 2016; Planchard et al. 2018; Møller et al. 2021), and the recent study in which publicly available Illumina MiSeq data were used to perform de novo assembly of the Arabidopsis Col-0 mitochondrial genome, followed by additional verifications against other experimental DNA sequence data (Sloan et al. 2018). There are 32 unique mitochondrial-encoded protein sequences and 3 tentative proteins encoded by ORFs (Supplemental Data Set 2). Complete protein-coding transcripts for mitochondrial-encoded Ndh1,2,5 are each generated by trans-splicing of individual exons. Each of these exons has ATMG identifiers in Araport11/TAIR10 but these 3 proteins should each be represented by only 1 ATMG identifier; for simplicity, we selected the identifiers with the lowest number (Supplemental Data Set 2). These 32 mitochondrial-encoded proteins are 9 subunits of the NAD complex (complex I), 1 subunit of complex III, 3 subunits of complex IV, 6 subunits of the ATP−synthase complex (complex V), 5 proteins involved in cytochrome biogenesis, 7 ribosomal proteins, and a maturase. In addition, there are the tentative proteins OrfX/TatC/MttB (candidate transporter protein), Orf114, and Orf240A. All mitochondrial protein-coding RNAs, except Cox1 and Orf114, are predicted to undergo RNA editing that leads to 1 or more amino acid changes. There are a total of 420 predicted editing sites that would result in a change in amino acid. Based on experimental RNA data, 57 of these editing sites are considered low-frequency sites (Supplemental Data Set 2); essentially, this means when sequencing cDNA libraries, these mRNAs are mostly found in their unedited form. These edit sites are based on information from Bentolila, Oh, et al. (2013) and Sloan et al. (2018) and other publications, as well as from communications with the scientific community with expertise in plant mitochondrial editing (see Acknowledgments).

Detection of the predicted unedited and edited plastid proteoforms

We detected 63 plastid-encoded proteins at the most confident (canonical) level, as well as 12 proteins at lower confidence levels (tiers), whereas 3 proteins (NdhG/Ndh6, PsbM, and Rps16) and ORF Ycf15 were not detected, i.e. there were no matched MS/MS spectra (PSMs) that passed our confidence threshold. The lower confidence level identifications fall in 2 tiers, i.e. insufficient evidence (1 or more uniquely mapping peptides, but none reach 9 residues in length) or weak (at least 1 uniquely mapping peptide of ≥9 residues in length, but otherwise does not meet our criteria for canonical). The system of different confidence tiers was explained in detail in van Wijk et al. (2021). Protein sequence coverage (by mapped peptides) of these identified proteins ranged from 10% (Ycf10/CemA) to 100% (Supplemental Data Set 1). As expected, RbcL was identified with the highest number of PSMs (1.6 million), followed by Cf1β and Cf1α with 0.5 and 0.3 million PSMs, respectively (Supplemental Data Set 1). The 3 undetected proteins (NdhG/Ndh6, PsbM, and Rps16) and the proteins assigned

"weak" identifications are either very hydrophobic (gravy indices > + 1) or small (<100 aa).

Rps16 is a 79 aa basic protein (isoelectric point 10.57) due the presence of a high number of K and R residues; such high basicity is quite common for ribosomal proteins. We were surprised by its lack of identification because tryptic digestion of the protein should result in 2 tryptic peptides of 7 aa residues (and several additional peptides when allowing for 1 missed cleavage). It is therefore possible that Rps16 is a pseudogene in Arabidopsis, but we cannot be entirely conclusive because this is based on negative evidence (i.e. lack of observed tryptic or semitryptic peptide(s)). Widespread pseudogenization of *rps16* in the angiosperm chloroplast genomes via the loss of its splicing capacity was described, even when the *rps16* encoded in the chloroplast genome is transcriptionally active (Ueda et al. 2008; Roy et al. 2010). It was suggested that the chloroplast-encoded Rps16 in mono- and dicotyledonous plants has been substituted by the product of nuclear-encoded Rps16, which was transferred from the mitochondria to the nucleus before the early divergence of angiosperms (Ueda et al. 2008; Roy et al. 2010). Arabidopsis Col-0 has 2 nuclear-encoded Rps16 homologs (AT4G34621 and AT5G56940; RPS16-1 and RPS16-2, respectively) of which AT4G34621 was shown to be an essential gene (Tsugeki et al. 1996). Based on confocal microscopy with fluorescent reporter constructs, it was suggested that the Rps16 homologs have dual-localization signals for plastids and mitochondria (Ueda et al. 2008). Arabidopsis PeptideAtlas and PPDB show that RPS16-1 accumulates at much higher levels than RPS16-2, and inspection of the metadata in PeptideAtlas suggests that RPS16-1 localizes to chloroplasts whereas RPS16-2 localizes to mitochondria (note—we use capital letters for nuclear-encoded proteins as per convention for Arabidopsis, but bacterial annotation convention for organelle-encoded proteins as established over the last decades).

The 176 aa NdhG/Ndh6 (gravy 1.17) was not identified because it has 5 predicted transmembrane domains and only 1 lysine (K128) and 1 arginine (R175). Consequently, trypsin digestion does not result in any peptides suitable for detection by MS. It was shown that its translation is under control of nuclear-encoded PROTON GRADIENT REGULATION 3 (PGR3), and it is not a pseudogene (Rojas et al. 2018; Higashi et al. 2021). PsbM is a 34 aa protein with 1 transmembrane domain and contains only 1 lysine (K28) and no arginine. Trypsin digestion will result in a 28 aa N-terminal peptide spanning the transmembrane domain and a 6 amino acid long C-terminal peptide that is below the minimal length for protein identification; this explains the lack of observation. PsbM is located in the stroma lamellae and is involved in biogenesis of PSII, but it is not a PSII subunit itself (Plochinger et al. 2016).

Chloroplast RNA editing (or lack of editing) at the protein level may be detected in the form of (un)edited peptides; these peptides must have at least 7 amino acids and a decent propensity to be ionized in the source of the mass spectrometer. In total, after manual validation, we observed peptides covering 10 of the 32 editing sites across 8 plastid-encoded proteins (Table 1). The level of experimental support for these editing sites is determined by the number of distinct peptides and the number of observations (PSMs). For the highly observed sites, 72P-S in ATCG00580.1 (Cytb559α) and 30S-L in ATCG00840.1 (Rpl23A), these unedited sites were only at 0.2% and 2% frequency, respectively. The RpoC1 editing site (163S-L) was detected with only 8 PSMs and only 1 distinct peptide, but none showed the edited form. This is consistent with the relatively low levels of RNA editing (24%) observed at the transcript level in wild-type Arabidopsis plants (Chateigner-Boutin et al. 2010; Bentolila, Oh, et al. 2013). For the other 6 editing sites, only the edited form was detected. Figure 2 displays some representative spectra for both the unedited and edited forms for these 2 cases (Cytb559α and Rpl23A). These 4 PSMs are of extremely high quality and provide unequivocal evidence of both edited and unedited forms. To evaluate the possible biological importance for these partial edits, we looked at the metadata for particular features of the samples (e.g. nonphotosynthetic tissue). However, no obvious trends were found, and we hypothesize that some low level of missed editing is inevitable in plastids and it is not detrimental to their overall successful function.

Detection of the predicted unedited and edited mitochondrial proteoforms

We detected all 32 mitochondrial-encoded proteins with 27 proteins at the most confident (canonical) level and 5 at lower confidence levels, but we did not detect proteins encoded by the 3 ORFs (Supplemental Data Set 2). Sequence coverage ranged from 11% (Ccb203/Ccmfn2/Ccb6n2) to 100% (ATP1), and matched PSMs ranged from 2 (Ccb206/CcmB/Ccb2) to ~0.2 million PSMs (ATP1). The 5 proteins identified at lower confidence levels are the pair of Atp6-1,2 homologs (identical in their C-terminal 252 aa) with 2 peptides (DNVPFLQSIA and ELDHTPAELLGSK) that were able to distinguish between Atp6-1 and Atp6-2 (Supplemental Data Set 2), Ndh4L, Nad3, and Ccb203/Ccmfn2/Ccb6n2. There were no PSMs that mapped to the proteins predicted to be encoded by the 3 ORFs (Orf114, Orf240A, and OrfX/TatC/MttB). If translated, these 3 ORF proteins are of medium size (114, 240, and 206 aa) and not particularly hydrophobic (gravy index 0.078, 0.435, and 0.39) with ample tryptic cleavage sites, and thus, their physicochemical properties do not interfere with detection by MS/MS. Therefore, the complete lack of observation (PSMs) suggests that these 3 ORFs are in fact pseudogenes; however, as discussed below, there is evidence that *tatc* transcript and protein do accumulate (Carrie et al. 2016; Planchard et al. 2018).

The mitochondrial intron maturase 2C type II (MatR; ATMG00520.1 and P93307) is annotated in several sources as having a non-AUG start site. As shown in Fig. 3A, sequences from UniProtKB (P9907), TAIR10 (TAIR10_ATMG00520.1),

Table 1. Verified protein editing status of 10 editing sites in 8 plastid-encoded proteins based manual evaluation of the second release of the Arabidopsis PeptideAtlas

PeptideAtlas-ATCG identifier[f]	Name	Edit site	Unedited PSMs[a]	Edited PSMs[a]	% edited based on MS/MS	% edited based on RNA-seq[b]	Most common unedited peptide sequence and PSMs[c]	Most common edited peptide sequence and PSMs[d]
ATCG00330.1	Rps14	50-P-L	0	94	100	89		IHGKLQSLPR (94)
ATCG00500.1	AccD	265-S-L	0	185	100	99		IELLIDPGTWNPMDEDMVSADPIK (185)
ATCG00570.1	PsbF/ Cytb559b	26-S-F	0	212	100	99		WLAVHGLAVPTVFFLGSISAMQFIQR (212)
ATCG00580.1	PsbE/ Cytb559a	72-P-S	26	15,015	100	100	FDPLEQLDEFSR (26)	FDSLEQLDEFSR (12,794)
ATCG00670.1	ClpP1	187-H-Y	0	469	100	90		DVFMSATEAQAYGIVDLVAVQ (469)
ATCG00740.1	RpoA	67-S-F	0	58	100	84		FENIPHDYSNIAGIQESVHEILMNLNEIVLR (58)[e]
ATCG00180.1	RpoC1	163S-L	12	nd[e]	nd[e]	nd	GSFEYEIQSWK (12)	
ATCG00840.1	Rpl23-1,2	30-S-L	33	1,609	98	84	NQYTFNVESGSTR (33)	NQYTFNVESGLTR (1,529)
ATCG00890.1	NdhB-1,2/ Ndh2-1,2	249-S-F	0	1	100	99		LSLAPFHQWTPDVYEGSPTPVVAFLSVTSK (1)
ATCG00890.1	NdhB-1,2/ Ndh2-1,2	277-S-L	0	28	100	98		VAALALATR (28)
ATCG00890.1	NdhB-1,2/ Ndh2-1,2	279-S-L	0	28	100	96		VAALALATR (28)

[a]Manually verified.
[b]From Bentolila, Oh, et al. (2013).
[c]Unedited residue in blue.
[d]Unedited residue in red.
[e]We only searched the unedited site, as we did not include the edited variant in the search space.
[f]Within the PeptideAtlas database, these protein IDs are listed as PeptideAtlas_ATCGxxxxx.1—as to differentiate them from ATCG numbers in TAIR.

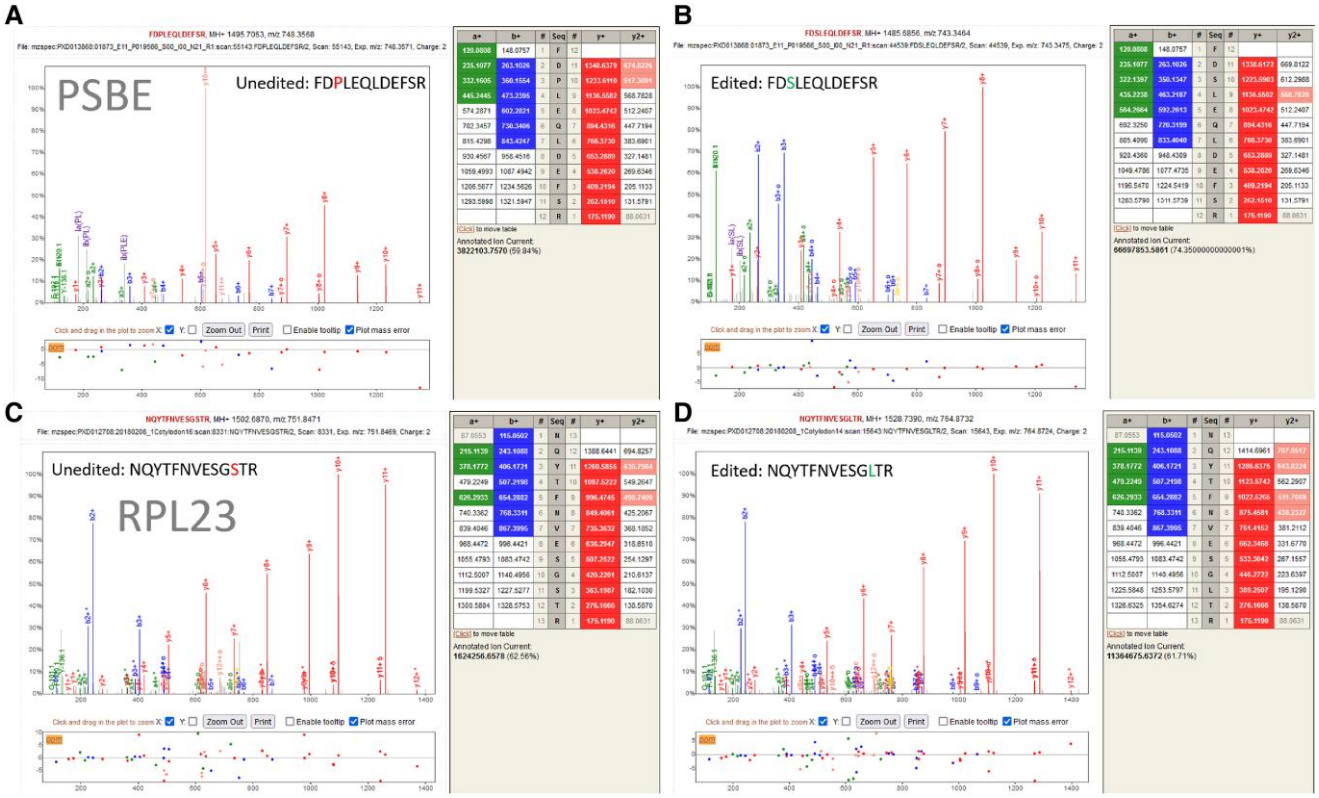
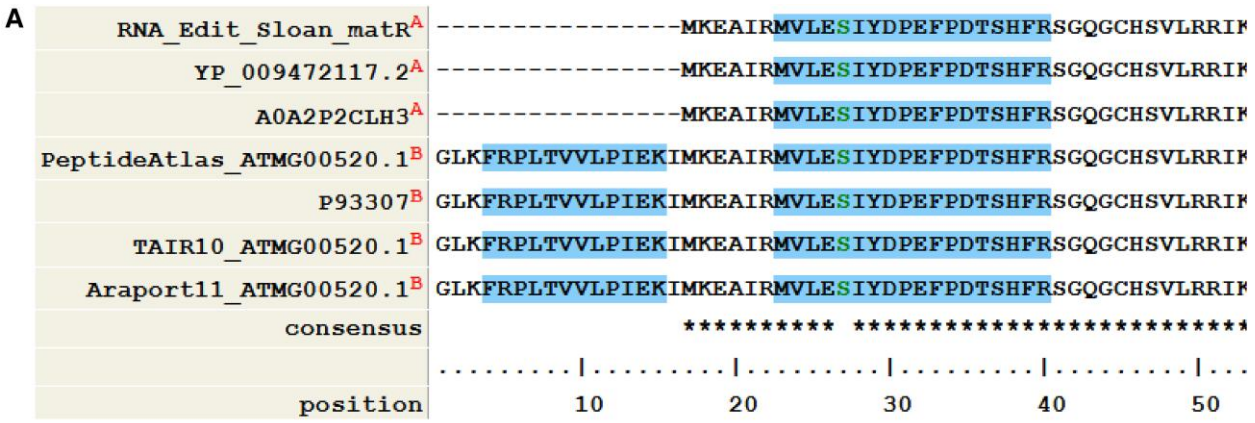
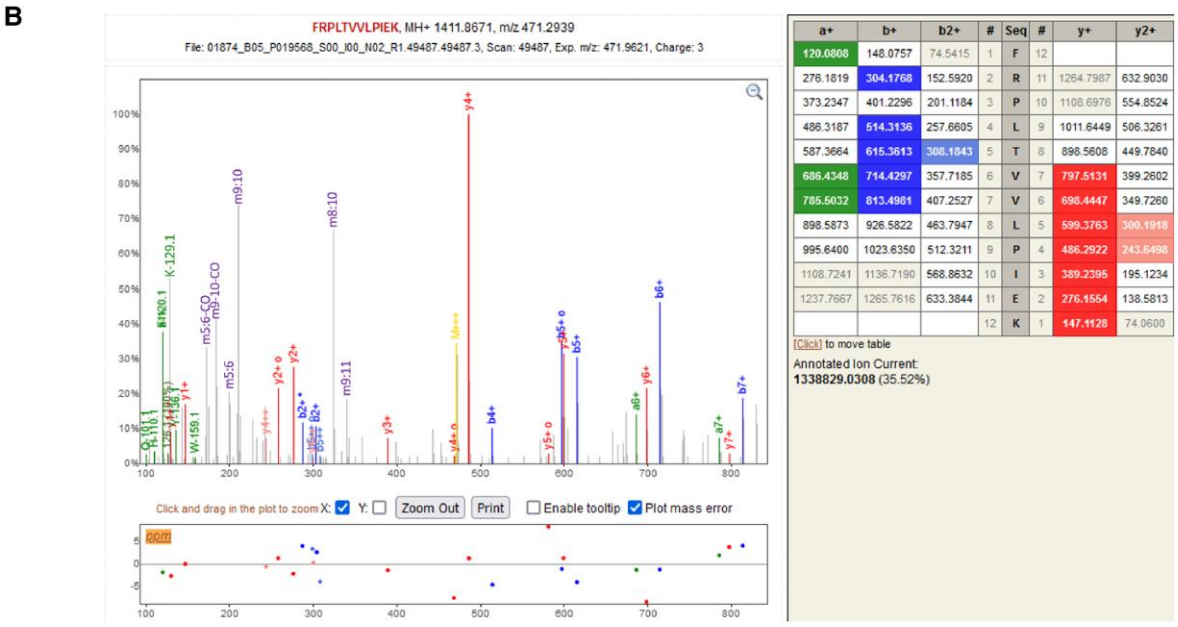


Figure 2. Representative MS/MS spectra for positively charged peptides of unedited and edited forms of Cytb599α (PsbE) and Rpl23-1,2. **A)** The unedited form FDPLEQLDEFSR of 72P/S of Cytb599α. Note the very strong y10++ ion indicative of a proline. **B)** The edited form of FDSLEQLDEFSR of 72P/S of Cytb599α with a visible but muted y10++ ion. **C)** The unedited form NQYTFNVESGSTR of 30S/L of Rpl23. Detection of nearly all y-ions provides ample evidence of the unedited form. **D)** The edited form NQYTFNVESGSTR of 30S/L of Rpl23 with an equivalently perfect ladder of y-ions. All panels show the MS/MS spectrum of a specific peptide with annotated ion peaks (red, blue, or green), internal fragments (gray), and unassigned peaks (gray). Red annotations indicate the y-ions (i.e. peptides that include the C-terminal residue of the peptide), the b-ions (i.e. peptides that include the N-terminal residue of the peptide), and a-ions (i.e. peptides that include the N-terminal residue of the peptide but involved the cleavage of a C–C bond rather than the peptidyl bond). The panel below each spectrum shows the relative mass errors (in ppm) for the assigned and unassigned MS/MS ion peaks following the color coding of the spectrum. The right-hand side table in each panel shows the masses of the a-, b-, and y-ions and the amino acid residues of the peptides. Masses are color coded if they are assigned in the MS/MS spectrum.

Araport11 (Araport11_ATMG00520.1), and the PeptideAtlas consensus (PeptideAtlas_ATMG00520.1) all start with a glycine as the N-terminal residue. Sloan et al. (2018) annotated a start site at the first methionine 17 residues further downstream, represented in UniProtKB with A0A2P2CLH3 and RefSeq as YP_009472117.2. UniProtKB cites Sloan et al. (2018) for this entry. Obtaining MS evidence of the methionine start site is difficult on account of the closely spaced lysine and arginine residues that follow the methionine, and that part of the sequence is indeed not represented in PeptideAtlas (Fig. 2A). There is 1 tryptic peptide detected in PeptideAtlas (FRPLTVVLPIEK) that appears to support an N-terminus that is more upstream than suggested in Sloan et al. This tryptic peptide (FRPLTVVLPIEK) is supported by 6 PSMs that pass our threshold in 4 separate samples from 2 data sets (PXD010324 and PXD013868). Most of these spectra have a signal-to-noise ratio less than 100 and are not very impressive, but 1 higher quality spectra shown in Fig. 3B appears to provide quite compelling evidence for the detection of FRPLTVVLPIEK. This spectrum (USI mzspec:PXD013868:01874_B05_P019568_S00_I00_N02_R1:scan:49487:FRPLTVVLPIEK/3) has nearly all major peaks explained, albeit with substantial internal fragmentation. This peptide does not map to any other entries in our comprehensive database, even considering I/L substitution. There is growing evidence that many proteins sometimes or often begin translation at a non-AUG start site (Cao and Slavoff 2020). However, glycine is an unusual start residue, whereas valine (GUG) and leucine (UUG) are the most common after methionine (AUG; Belinky et al. 2017). It is therefore possible that the leucine following the glycine is the true start site (hence, N-terminus would be LKFRPFRPLTVVLPIEK). Alternatively, as discussed for Rps19, UUG can serve as the initiation codon for formylmethionine via wobble pairing with tRNA-*f*Met (CAU); a recent paper in tobacco showed that GUG and UUG can be start codons (Sadhu et al. 2023). This would result in the N-terminus MLKFRPFRPLTVVLPIEK.





a+	b+	b2+	#	Seq	#	y+	y2+
120.0808	148.0757	74.5415	1	F	12		
276.1819	304.1768	152.5920	2	R	11	1264.7987	632.9030
373.2347	401.2296	201.1184	3	P	10	1108.6976	554.8524
486.3187	514.3136	257.6605	4	L	9	1011.6449	506.3261
587.3664	615.3613	308.1843	5	T	8	898.5608	449.7840
686.4348	714.4297	357.7185	6	V	7	797.5131	399.2602
785.5032	813.4981	407.2527	7	V	6	698.4447	349.7260
898.5873	926.5822	463.7947	8	L	5	599.3763	300.1918
995.6400	1023.6350	512.3211	9	P	4	486.2922	243.6498
1108.7241	1136.7190	568.8632	10	I	3	389.2395	195.1234
1237.7667	1265.7616	633.3844	11	E	2	276.1554	138.5813
			12	K	1	147.1128	74.0600

Figure 3. A sequence alignment of mitochondrial MatR in several different sources and spectral matches. **A)** UniProtKB's primary entry, TAIR10, and Araport11 all maintain an early non-AUG start site, while UniProtKB's alternative entry A0A2P2CLH3 and RefSeq have a later start site at the first methionine, citing Sloan et al. (2018). Sequence with blue shading is detected in PeptideAtlas, while sequence with a white background is not detected. The serine that is indicated in green at position 27 is a subject of a potential variant, not detected in PeptideAtlas. The N-terminal glycine residue is encoded by UUG. Importantly, UUG might serve as an initiation codon for formylmethionine via wobble pairing with tRNA-*f*Met (CAU) resulting in the N-terminus MLKFRPFRPLTVVLPIEK. **B)** mzspec:PXD013868:01874_B05_P019568_S00_I00_N02_R1:scan:49487:FRPLTVVLPIEK/3 using the Lorikeet spectrum viewer that shows compelling evidence for the detection of FRPLTVVLPIEK, a peptide that only maps to a potential early start site of MatR. The spectrum may be examined at http://proteomecentral.proteomexchange.org/usi/ using the above USI. The internal fragmentation ions (e.g. m9:10 indicates a b-type ion of residues 9 and 10, i.e. PI) are not natively labeled in the spectrum viewer (Lorikeet) and added manually above.

There are in total 420 predicted nonsynonymous edits across the mitochondrial-encoded proteins. Based on experimental RNA data (Bentolila, Oh, et al. 2013), 57 editing sites are considered low-frequency sites (Supplemental Data Set 2). In total, we observed 157 of the editing sites (Supplemental Data Sets 2 and 3). We applied the same criteria and manual evaluation of low-frequency events as for the plastid-encoded proteins. Manual evaluation of matched spectra uncovered several incorrect assignments for the S/L editing site with assignment of formylated S instead of L, the edited form. This was due to a software assignment error in selection of the correct isotope, which should be the monoisotopic form of the selected precursor ion (i.e. all carbon atoms are ^{12}C), but instead, the precursor ion with 1 ^{13}C was selected. Figure 4 shows an example of a peptide for Nad7 that appears to provide evidence of unedited protein, but the explanation provided by the edited form is better, and the most likely explanation is that the instrument-selected precursor mass belongs to a cofragmented contamination precursor, rather than the proposed explanation. This demonstrates one of the challenges in making unedited/edited

assignments, and why manual inspection of some assignments is required.

A summary of the validated editing site observations for mitochondrial-encoded proteins is provided in Supplemental Data Sets 2 and 3 organized by specific editing site. Figure 5 displays these results for the 26 proteins with at least 1 editing site observed with at least 5 PSMs. Almost all sites have nearly complete edits at 95% to 100% or nonedits 0% to 10%, with the single exception of Rpl5, with an edited/unedited ratio of 0.6. In the case of 1 minor frequency site based on RNA (ATMG00830 124S/L with 5/0 edited/unedited), we observed high frequency at the peptide level. Importantly, all other 20 edit sites that we observed with low editing frequency were indeed assigned as low-frequency sites at the RNA level (Sloan et al. 2018). Conversely, all sites that we observed to be edited at high frequency were also assigned as high-frequency editing sites at the RNA level.

Frequencies and observations of the different types of edits at the amino acid level

RNA editing across plastids and mitochondria leads to 14 possible amino acid changes (accounting for these 452 total possible edits) including 2 types of RNA edits (Q-> stop, R-> stop) that lead to introduction of a stop codon (6 and 4 sites, respectively; Table 2). The predicted S to L edit is the most frequent amino acid change (117 times), whereas the T to M edit is the least frequent (in just 5 cases). Across plastid and mitochondrial proteins, we observed all types of edits at the peptide level with the exception of the P to F change and the stop codon introductions (Table 2). Evaluation of an introduction of an additional stop codon is challenging as this requires the identification of a diagnostic C-terminal peptide (and these stops would lead to truncation, rather than extension). We did not observe peptides that support new stop codons that result in such truncations. In 166 out of these 452 cases, we did obtain peptide coverage of the editing site, such that we could potentially determine if an edit did indeed occur (Table 2). Because sequence identification by MS/MS is based on the predicted mass for each residue, amino acid changes are detected as a change in mass. To evaluate possible misassignments (or edits and PTMs), we calculated the predicted delta mass for each of the 12 changes (Table 2) and they range from 10.02073 to 60.03638 Da, which is well within the mass accuracy for modern mass spectrometers. Only in the case of the S-L and P-L edits does the delta mass have some similarity to a common posttranslational modification (PTM). The S-L transition corresponds to 26.05203 Da, which is somewhat close (delta is 1.94288 Da) to formylation (27.99491 Da), and the P-L transition corresponds to 16.0313 Da, which is somewhat close (delta is 0.03639 Da) to oxidation (15.99491 Da). However, also these are easily distinguishable with the mass accuracy of current mass spectrometers and should not result in false positives. However, as we discussed above and in Fig. 3, a few cases of misassignment of S to formylated S, rather than S to L, were observed due to simultaneous fragmentation of 2 peptides and incorrect selection of the monoisotopic peptide peak.

One other consideration that might lead to biases in detection is the change in physicochemical properties of the amino acid of the unedited and edited position, possibly biasing the detectability by MS/MS. In particular, the transition from R to W, R to C, and H to Y results in the loss of positive charge (at the acidic pH used in typical liquid chromatography–MS workflows) and increase in hydrophobicity, as well the loss of a tryptic cleavage site (since trypsin cleaves after R and K). The loss of the tryptic cleavage site can however be beneficial for improved coverage of the predicted edit site as a larger peptide is generated. We mapped a substantial portion (25% to 55%) for all types of RNA editing sites, with the exception of the 2 P-F transitions (Table 2). Therefore, the editing frequencies of plastid and mitochondrial proteins appeared unbiased for editing frequencies for the different types. Finally, the MS/MS results mapped 26 out of 56 RNA-seq-based minor-frequency sites (46%) compared to 140 out of 395 RNA-seq-based high-frequency sites (35%), indicating that the peptide analysis was unbiased for its assessment of low- (minor) vs high-frequency edits.

PTMs of the plastid- and mitochondrial-encoded proteomes

Like nuclear-encoded proteins, the organelle-encoded proteins can undergo several types of in planta PTMs. We specifically searched for N-terminal Acetylation (NTA) for all proteins, and for lysine acetylation, phosphorylation (lysine, threonine, and tyrosine) and lysine ubiquitination in the case of experimental data sets that were deliberately affinity enriched for these PTMs (van Wijk, Leppert, et al. 2023). A sophisticated PTM viewer in PeptideAtlas allows detailed examination of these PTMs, including direct links to all spectral matches, and we recommend using the PeptideAtlas to evaluate specific PTM sites if these are of particular interest to the reader. PTM identification rates strongly depend on the confidence level (minimal probability threshold) of PTM assignment. Here we used localization probability $P \geq 0.95$ from PTMProphet (Shteynberg et al. 2019) for each PTM and also required at least 3 PSMs for a PTM at a specific residue to be included in the summary (Supplemental Data Set 4 for the data on all 4 PTMs). In general, higher numbers of repeat observations (PSMs) for a specific PTM at a residue improve the reliability of the assignment. Conversely and importantly, peptides with high PSM counts (e.g. hundreds or more) for which the vast majority (e.g. >90%) of peptides do not have a reported PTM at $P > 0.95$, are possibly false PTM discoveries. We evaluated the results for false positives and possible pitfalls in various ways, including spot checking matched spectra and proteins to which PTMs were mapped. Supplemental Data Set 4 provides these details and also provides direct links to peptide and spectral URLs (within the PeptideAtlas DB).

The chloroplast acetylation machinery in Arabidopsis consists of 8 General Control Non-repressible 5 (GCN5)–related

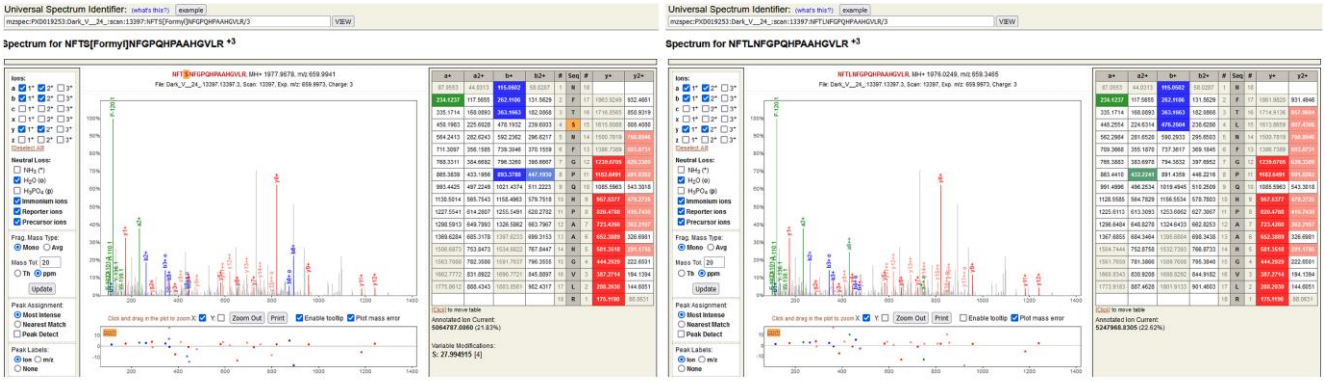


Figure 4. Example of likely unedited/edited conflation in the case of Nad7. The left panel shows the search engine result, which matches the spectrum to an unedited peptide, but with the uncommon formylation mass modification (+27.9949) on the unedited serine. The precursor *m*/*z* delta is only 5 ppm, which is within our search tolerance. Although some unlabeled gray peaks can be explained by internal fragmentation and imperfect annotation, major unexplained peaks at 732 and 892 are clearly contamination by another cofragmented precursor. The b8 annotation is spurious as it is on the wrong isotope and completely unexpected on the wrong side of a proline. The right panel shows a more likely explanation, after manual inspection. The S[Formyl] is replaced by the expected editing result, an L. This allows both the b+ and y++ series to extend further in both directions. The spurious b8 in the left explanation disappears. However, the precursor *m*/*z* is now way off at 1.9524 *m*/*z*, well outside our search tolerance. Attempts to rectify this with deamidation do not improve the fragmentation explanations. The most likely explanation is that the instrument-selected precursor mass belongs to the contamination precursor, rather than the proposed explanation. This demonstrates an example of potential hazards in making unedited/edited assignments, and why manual inspection of some assignments is required. The spectrum and potential explanations can be explored by the reader using the provided USIs via the ProteomeXchange USI page at http://proteomecentral.proteomexchange.org/usi/. All panels show the MS/MS spectrum of a specific peptide with annotated ion peaks (red, blue, or green), internal fragments (gray), and unassigned peaks (gray). Red annotations indicate the *y*-ions (i.e. peptides that include the C-terminal residue of the peptide), the b-ions (i.e. peptides that include the N-terminal residue of the peptide), and a-ions (i.e. peptides that include the N-terminal residue of the peptide but involved the cleavage of a C–C bond rather than the peptidyl bond). The panel below each spectrum shows the relative mass errors (in ppm) for the assigned and unassigned MS/MS ion peaks following the color coding of the spectrum. The right-hand side table in each panel shows the masses of the a-, b-, and y-ions and the amino acid residues of the peptides. Masses are color coded if they are assigned in the MS/MS spectrum.

N-acetyltransferase (GNAT) enzymes that catalyze both N-terminal and lysine acetylation of proteins, whereas no mitochondrial localized protein acetyltransferases are known (Pozoga et al. 2022). NTA at position 1 (the start methionine) or 2 was observed for 17 plastid proteins but not for any of the mitochondrial-encoded proteins (Supplemental Data Set 4). This is consistent with previous reports that chloroplast-encoded proteins undergo frequent NTA (here observed at residues A, S, T, V, I, and R; Zybailov et al. 2008; Dinh et al. 2015; Rowland et al. 2015; Bienvenut et al. 2020; Willems et al. 2021) but mitochondrial proteins do not (Huang et al. 2009). For a few of these chloroplast-encoded proteins, we observed NTA at both positions 1 and 2 (PsaA, Cf1α, Cf1β, and D1), but NTA at position 2 always had far more PSMs. For Cf1α, Cf1α, RbcL, and Cp47, the search also identified NTAs much farther downstream of the N-termini. However, in nearly all cases, these downstream PTM sites likely represent false positives since most of these peptides were observed without NTA (i.e. the frequency of assigned NTA for these observed peptides was very low). We observed lysine side-chain acetylation of only 2 mitochondrial proteins (ATP1 and NAD1; 4 sites), but 15 chloroplast proteins belonging to the ATP–synthase (Cf1α,β and Cfo-I,III), the cytb6f complex (Cytf and Cytb6), Photosystem I (PsaA,B), Photosystem II (CP43, CP47, D2, and PsbH), Tic214, and RbcL (66 sites across these 15 proteins). The number of PSMs per K-acetyl site that passed the *P*-value thresholds range from 3 to 159; a higher number of PSMs should associate with higher overall confidence of significance of the observation. The ratio between PSMs that include a PTM and total PSMs for peptides that include that site can reflect the false discovery rate (FDR) of PTM discovery, and PTM sites with low ratios should be treated with reservation.

We observed phosphorylation of 12 plastid-encoded proteins and just 1 mitochondrial protein (RPS3, S-112; Supplemental Data Set 4). The phosphorylated plastid proteins include RbcL, ribosomal Rsp7/14, Ycf1, PsaA, 3 Photosystem II core proteins (Cp47, PsbH, and PsbL), and 4 ATP–synthase subunits. The number of PSMs for most of these p-sites was relatively low, especially compared to the high number of PSMs for these peptides from nonphosphorylation data sets (which were not searched for p-peptides). p-sites with only few PSMs are therefore very rare events especially considering the large number of spectra searched; however, several p-sites (RbcL-330T and Rps7-S93) were also observed in prior, carefully evaluated data sets (Schonberg et al. 2017).

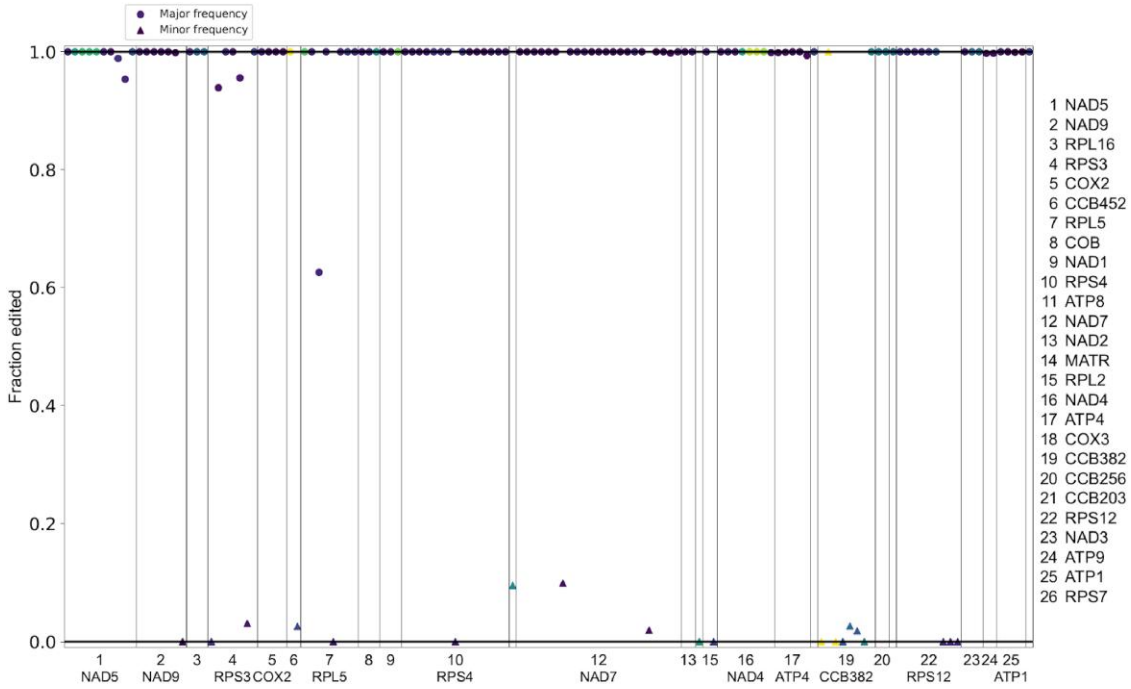


Figure 5. Overview of mitochondrial editing sites based on Supplemental Data Set 3 for the 26 proteins with at least 1 editing site with at least 5 observations. High-frequency edits (based on RNA) are depicted with circles and low-frequency edits (based on RNA) with triangles. Vertical lines separate the proteins, which are numbered at the bottom and labeled on the right. Observed sites at higher frequencies are darker in color, while sites with low numbers of observations are shown with a lighter color.

Based on the 2 large-scale ubiquitination studies included in build 2 (Walton et al. 2016; Grubb et al. 2021), we observed only ubiquitination of chloroplast Rps7 (note that we observed ubiquitination for more than 1,000 nuclear-encoded proteins; van Wijk, Leppert, et al. 2023). This ubiquitination site in Rps7 (K13) was only observed in the data from Walton et al. (2016) and with only 3 PSMs; this likely represents an unintended alkylation, which results in an identical mass. The lack of ubiquitination for organelle-encoded proteins is consistent with our reanalysis of a very recent ubiquitination study that was otherwise not included in the second Arabidopsis PeptideAtlas release (van Wijk, Leppert, et al. 2023).

Discussion

Consensus sequences, identification, and coverage of the plastid- and mitochondrial-encoded proteomes

Originally, sources reported around 88 protein-coding plastid genes and 31 to 33 protein-coding mitochondrial genes in Arabidopsis (Unseld et al. 1997; Sato et al. 1999; Sloan et al. 2018). Unfortunately, several of the last iterations of the Arabidopsis community genome annotations (including TAIR10 and Araport11) include 122 mitochondrial-encoded protein identifiers (ATMG). It is not entirely clear why these additional identifiers and sequences were introduced. Furthermore, both plastid and mitochondrial cDNAs are generated by trans-splicing; at the protein level, these individual exons are not stable protein products, and the proteins from multiple spliced transcripts should be represented by a single protein identifier. Finally, the plastid genome includes 2 IRs mostly encoding identical protein products (the exception is the nonfunctional and truncated ATCG01000). To better determine the accumulation of the plastid and mitochondrial proteins, it is therefore imperative that a nonredundant set of identifiers of the full-length protein sequences is available. With help of several experts in the international research community, this study assembled this nonredundant set of proteins and also evaluated several conflicting start and stop sites.

Based on the systematic reprocessing and database search of >250 million raw MS/MS spectra as part of the recent release of Arabidopsis PeptideAtlas, we confirmed accumulation of 75 plastid-encoded proteins but not for plastid-encoded Rps16, PsbM, NdhG/Ndh6, and Ycf15/Orf77. After evaluation of additional information, we suggest that plastids Rps16 and Ycf15/Orf77 are likely pseudogenes and that the plastid genome encodes 77 stable proteins, some of which are very hard to detect by MS due to their small size and hydrophobic nature. Based on the recent Arabidopsis PeptideAtlas release, we also confirmed accumulation of 32 mitochondrial-encoded proteins, but no protein products for Orf240, Orf114, and Orfx/Tatc/Mttb were detected. We conclude that the mitochondrial genome encodes 32 stable proteins and, based on immunoblotting in Carrie et al. (2016), that TatC might accumulate at low levels.

Table 2. Summary of consensus editing types at the amino acid level (i.e. consequences of edits) and their frequencies across plastid- and mitochondrial-encoded proteins

Unedited–edited	No. of searched edit sites (major and minor)	No. of detected major sites (edited or unedited)	No. of detected minor edit sites (edited or unedited)	% detected	No. of PSMs unedited major	No. of PSMs edited major	No. of PSMs unedited minor	No. of PSMs edited minor	Mass of unedited aa (Da)	Mass of edited aa (Da)	Delta mass (Da)
A-V	6	1	0	17	23	496	0	0	71.03711	99.06841	−28.0313
H-Y	27	12	2	52	110	8,974	96	2	137.05891	163.06333	−26.00442
L-F	21	3	4	33	0	2,484	687	1	113.08406	147.06841	−33.98435
P-F	8	0	0	0	0	0	0	0	97.05276	147.06841	−50.01565
P-L	92	29	2	34	24	34,190	222	0	97.05276	87.03203	10.02073
P-S	40	7	6	33	28	18,341	1,735	100	97.05276	113.08406	−16.0313
Q[a]	6	0	2	33	0	0	4,231	0	128.05858	n/a	128.05858
R[a]	4	0	1	25	0	0	207	0	156.10111	n/a	156.10111
R-C	30	7	1	27	0	1,820	37	1	156.10111	103.00919	53.09192
R-W	30	8	1	30	0	4,775	4	0	156.10111	186.07931	−29.9782
S-F	55	16	3	35	6	10,813	509	16	87.03203	147.06841	−60.03638
S-L	117	54	2	48	122	45,087	19	7	87.03203	113.08406	−26.05203
T-I	11	3	2	45	17	3,570	1,018	20	101.04768	113.08406	−12.03638
T-M	5	1	0	20	0	1	0	0	101.04768	131.04049	−29.99281
Total	452	141	26		330	130,551	8,765	147			

n/a, not available.
[a]The RNA edits result in a stop codon.

However, due to RNA editing, most of the mitochondrial stable proteins and 17 plastid-encoded proteins are modified as compared to the unedited sequence.

To provide the research community this updated set of protein identifiers and their sequence variants, we provide several sequence data sets that can be used to e.g. simply obtain the most likely protein sequence variant for stable proteins, and/or for small- or large-scale proteomics studies, as follows:

1) Protein IDs and unedited sequences (with the exception of essential edits for start and stop codons that need to be applied to generate a protein) for the 77 plastid-encoded and 32 mitochondrial-encoded proteins and predicted protein sequences of the 4 assigned pseudogenes and TatC.
2) Protein IDs and sequences after application of editing for the 77 plastid-encoded and 32 mitochondrial-encoded proteins and predicted protein sequences of the 4 assigned pseudogenes and TatC; minor-frequency edits are not applied but all high-frequency edits are included.
3) Proteins IDs and the >10,000 sequence variants (see Materials and methods) to allow for complete and exhaustive MS/MS database search of all possible edits and allow for partial editing (similar as we did in this study).
4) Protein IDs and their amino acid sequences with all possible variants supplied in PEFF format (https://www.psidev.info/peff). This PEFF format allows for encoding variants in the file. This is a Protein Standards Initiative format that some MS search engines (e.g. Comet) can handle. It is essentially like FASTA, but one can encode "at position N in the reference sequence, the amino acid can also be X" (this could also allow to include previously identified PTMs).

Finally, we anticipate that a consensus set of sequences and their editing variants will replace the current set of plastid and mitochondrial proteins in the next and upcoming Arabidopsis genome annotation (TAIR12; tinyurl.com/Athalianav12) and can also be downloaded from the Arabidopsis PeptideAtlas website.

Detection and abundance of plastid- and mitochondrial-encoded proteins

The number of PSMs observed per plastid- or mitochondrial-encoded protein is an indication of detection efficiency and abundance. Protein abundance is the consequence of transcription rate, translational efficiency, and protein turnover rate. Ribosome profiling, which determines the ribosome occupancy per gene, showed that mitochondrial Arabidopsis transcripts greatly differ in ribosome association levels (Planchard et al. 2018). In particular, mRNAs encoding c-type cytochrome maturation (CCM) proteins have particularly low ribosome densities when compared to mRNAs encoding respiratory chain subunits. The low translational activity of the 5 CCM genes is supported by the small number of PSMs observed (2 to 485) and compared to respiration chain subunits or ribosomal subunits (mostly >1,000 and up to ∼204,000; Supplemental Data Set 2). Two of the mitochondrial ORFs undetected in PeptideAtlas, Orf114 and Orf240A, were not revealed by ribosome profiling (Planchard et al. 2018), and considering their favorable predicted protein properties for detection by MS/MS, they

can therefore be dismissed as pseudogenes. The results concerning mitochondrial TatC ORF are conflicting as no PSMs were reported in PeptideAtlas, but a detection of translational activity by ribosome profiling may support its functionality (Planchard et al. 2018). Arabidopsis TatC lacks a classical start codon, and the predicted N-terminal residue is an asparagine; however, an antibody raised against a TatC-specific peptide was able to detect a faint band in isolated Arabidopsis mitochondria (Carrie et al. 2016). The binding was lost when the antibody was preincubated with the peptide, suggesting that the band was not an unspecific reaction. Furthermore, using immunoblotting and blue native-SDS-PAGE, TatC was found to colocalize with (nuclear-encoded) TATB in the same high molecular weight complex. A definitive answer as to whether TatC is present in the Arabidopsis mitochondrial proteome would require e.g. the immunoprecipitation of the protein by an antibody and analysis by MS/MS of the immunoprecipitated proteins. Our study might also have revealed a non-canonical codon start for MatR, as 4 PSMs were detected for a 16 aa peptide upstream of the methionine start codon. In addition, our study confirms the splice junctions for 11 of the 23 mitochondrial splicing events and 11 of the 15 plastid splicing events (Supplemental Figs. S3 and S4).

Impact of RNA editing on the plastid and mitochondrial proteomes

In this study, we performed systematic MS/MS-based analysis on the impact of RNA editing on the accumulated plastid and mitochondrial proteomes across many photosynthetic and nonphotosynthetic tissues, developmental stages, and (a)biotic conditions. Previously, only 2 different mitochondrial proteins (Atp6 and Rps12) encoded by partially edited transcripts were evaluated for the presence of peptides encoded by incompletely edited transcripts; it should be noted that this was in petunia (*Petunia hybrida*) and maize (*Zea mays*), not Arabidopsis (Lu and Hanson 1994; Lu et al. 1996; Phreaner et al. 1996). In the case of Atp6, protein from unedited transcripts was not detected despite the presence of partially edited transcripts on polysomes. Lu and Hanson (1994) took advantage of partial editing of a stop codon that shortens the protein by 13 aa, allowing production of an antibody against the unedited tail of Atp6. Antibodies raised to peptides specific to unedited portions of *rps12* transcripts could detect some accumulation of Rps12 encoded by incompletely edited transcripts (Lu and Hanson 1994, 1996; Lu et al. 1996; Phreaner et al. 1996; Williams et al. 1998). The small number of proteins analyzed so far has prevented definitive conclusions about the possible role of partial editing in generating protein polymorphisms, especially in the plant mitochondrion. In mammalian systems such as apolipoprotein B, C-to-U editing creates a stop codon so that 2 forms of the protein differing in size and translated from edited or unedited transcripts are produced with different function and tissue specificity (Davidson 1993). Similarly, differential RNA editing in the human brain can alter the properties of glutamate receptors (Seeburg and Hartner 2003). In contrast, the dogma in the plant organelle editing field is that RNA editing is a corrective mechanism allowing the production of functional proteins despite improper DNA sequences (Small et al. 2020, 2023). The importance of RNA editing for protein function is supported by the numerous editing mutants showing strong developmental defective phenotypes (Small et al. 2020). Furthermore, editing usually restores the occurrence of phylogenetically more conserved amino acids that are encoded by organisms that do not edit their RNA. However, the question of whether plant organelle editing results in the production of a pool of diverse proteins from differentially edited transcripts has never been experimentally addressed in an extensive way.

Figure 6 summarizes the quantitative relationship between the frequency of recorded plastid and mitochondrial editing sites detected at the level of RNA (by RNA-seq—from Bentolila, Oh, et al. 2013) and proteins (only listing sites detected with at least 5 PSMs). In the case of minor editing sites (assignment based on RNA-seq), these sites are also moderately edited at the protein level; hence for these proteins, the lack of editing does not prevent proteins from accumulating. However, Ccb382 124-S-L is a noticeable exception: whereas the editing extent is 9% at the transcript level, only edited proteins are observed for this site. For major frequency editing sites, editing at the RNA level was between 60% and 100%; for more than 30 of these sites, editing was incomplete and well below 100% (Fig. 6). However, editing at the protein level was nearly always 100% or close to 100%, except for RPS3 296-S-L that was edited at 94% (Fig. 6).

Among the 36 mitochondrial sites in which the editing extent is in the 50% to 90% range as measured by RNA-seq (Bentolila, Oh, et al. 2013), most sites (31) are carried by transcripts encoding ribosomal proteins (Supplemental Data Set 3—sites in bold). Except for Rpl5-22 and Rps3-296, all corresponding peptides carry only the edited version of the encoded amino acid (33 edited codons at a 100% occurrence). Given that ribosome profiling showed that most partially edited sites remain partially edited in polysome-associated mRNAs both in Arabidopsis mitochondria (Planchard et al. 2018) and in maize chloroplasts (Chotewutmontri and Barkan 2016), our results suggest that there must be some checkpoints either cotranslationally or posttranslationally that exclude the expression of partially edited transcripts. In maize chloroplasts, there are however 2 sites, *rpl2*-C2 and *ndhA*-C 563, that show a preferential translation of the edited RNAs. In the case of *rpl2*, the editing event creates an AUG start codon from an ACG precursor, while editing at the *ndhA* site is linked to the splicing of the group II intron in the *ndhA* pre-mRNA: the site is not edited in unspliced transcripts, and it is fully edited in spliced transcripts. The position of the intron between the edited site and the *cis*-element that specifies it prevents editing in unspliced transcripts. Translation that initiates on unspliced ndhA RNA would terminate at an in-frame stop codon within the intron. Thus, exon 2 is translated only from spliced

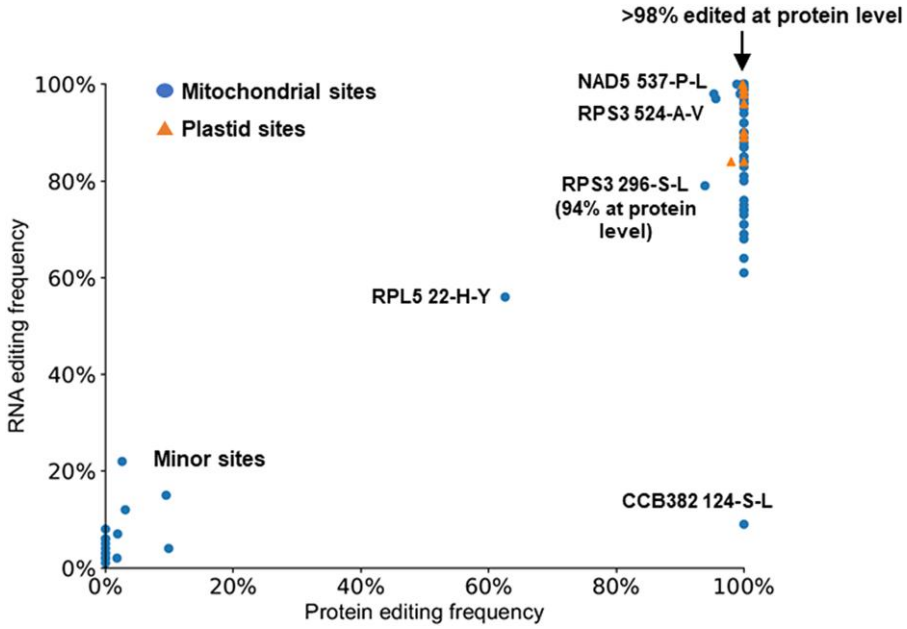


Figure 6. Quantitative relationship between editing at the RNA level (from RNA-seq; Bentolila, Oh, et al. 2013) and the protein level determined in this study (only sites detected with at least 5 PSMs are shown). The underlying data are from Table 1 and Supplemental Data Set 3.

RNAs, and these are 100% edited. Sampling with a reduced number of PSMs could explain the absence of unedited translated codons for a small number of the sites; 7 positions are covered by 1 to 61 PSMs (Supplemental Data Set 3). However, the likelihood of not detecting unedited peptides for the remaining 28 positions is very low given the relatively high number of PSMs (107 to 2837). This observation is consistent with the prior report of the absence of proteins corresponding to unedited *atp6* transcripts in petunia mitochondria (Lu and Hanson 1994). Our results also corroborate the observation reported for *clb19* mutants in Arabidopsis that show a complete lack of editing at 2 chloroplast sites located on *rpoA* and *clpP* transcripts (Chateigner-Boutin et al. 2008). The morphological phenotype of this mutant is caused by a dysfunction of the plastid-encoded polymerase (RpoA) supported by a decrease of the level of immunodetectable RpoA protein in *clb19* seedlings. The authors conclude that the lack of editing of *rpoA* transcripts has a critical effect on the translation or stability of the RpoA protein.

The only site in our study showing a substantial amount of partial editing both at the transcript level and at the protein level (56% of edited transcripts and 63% of edited proteins; Supplemental Data Set 3) is found in Rpl5 at position 22. Rps12 is the only other plant mitochondrial protein previously described in the literature that is found both in edited and unedited translated forms (Lu et al. 1996; Phreaner et al. 1996). In maize, immunological analysis with 2 antibodies specific for the unedited or edited version of a peptide covering 3 partially edited sites at positions 90, 95, and 97 demonstrated that both edited and unedited *rps12* translation products are present in mitochondria (Phreaner et al. 1996). However, only edited translation products accumulate in assembled maize mitochondrial ribosomes. The authors speculated that the unedited Rps12 protein might have taken a separate function from its role in translation as a component of the ribosome. As a rationale for that hypothesis, Rps12 from *Escherichia coli* is known to have nonspecific RNA binding protein activity and to facilitate intron splicing and RNA folding in vitro (Coetzee et al. 1994). In petunia, the same antibodies allow discrimination between the edited and unedited forms that result from the translation of 2 conserved editing sites with maize at positions 90 and 95. Unlike maize, RpsI2 proteins recognized by both edited and unedited specific antibodies are present in the petunia mitochondrial ribosomal fraction (Lu et al. 1996). The authors also showed that unedited *rps12* translation products were detected in plant species other than petunia (pumpkin [*Cucumis pepo*] and wheat [*Triticum aestivum*]), demonstrating that the polymorphism in mitochondrial *rps12* expression is widespread. The extent of editing in Arabidopsis measured by RNA-seq was 88% for both rps12-90 and rps12-95 sites, and the MS/MS analysis did only detect edited peptides (Supplemental Data Set 3).

In *rpl5*, the partially edited tyrosine codon located at position 22 in Arabidopsis (position 27 in the alignment shown in Fig. 7) is highly conserved. The tyrosine is genomically encoded in bacteria, green algae (*Nasuella olivacea*), nonvascular plants comprising liverworts (*Marchantia palacea*, *Calypogeia arguta*, and *Aneura pinguis*) and mosses (*Anomodon rugelii*, *Climacium dendroides*, *Physcomitrella patens*, and *Polytrichum commune*), and in some class of vascular plants, the lycopods, or clubmosses represented by *Isoetes engelmannii*. In angiosperms (*Oryza sativa*, *Silene latifolia*, and

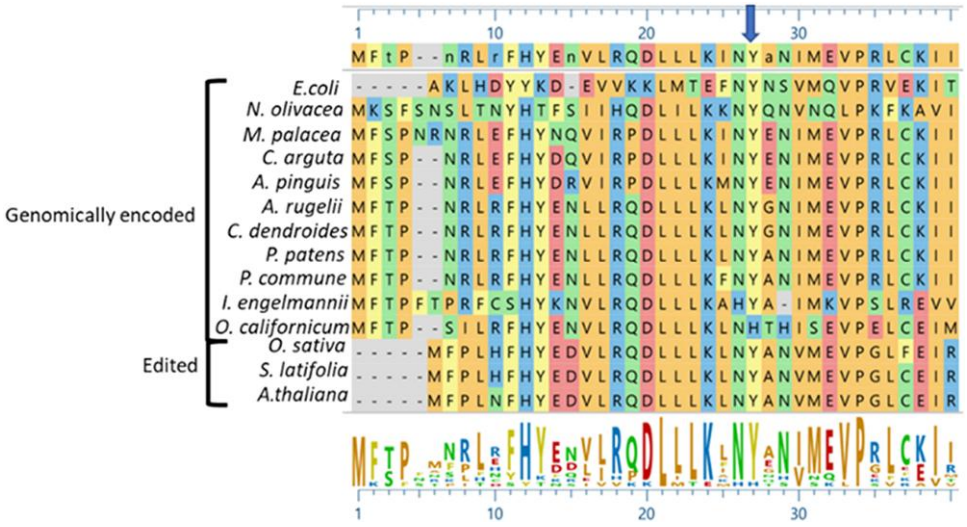


Figure 7. Alignment of the N-terminus of Rpl5 proteins in a wide range of species (bacteria, mammals, fungi, mosses, liverworts, ferns, kiwi, monocotyledons [*O. sativa*], and dicotyledons [*S. latifolia* and *A. thaliana*]). With the exception for the fern *O. californicum*, the conserved tyrosine that is either directly encoded or generated by RNA editing is indicated by an arrow.

A. thaliana), editing of the first nucleotide in the codon for histidine (CAC) results in the creation of the codon encoding tyrosine (UAC). In the fern *Ophioglossum californicum*, histidine is encoded in the mitochondrial *rpl5* gene as in the angiosperms, but evidence for editing at that position has not been established. Given the high level of conservation for tyrosine across different lineages, we can assume that its presence is required for a proper function of Rpl5 in the ribosome. Experiments in bacteria have shown that Rpl5 is a 5SrRNA-binding protein that is important for the incorporation of this ribosomal RNA into the 50S ribosomal subunit in vitro (Rohl and Nierhaus 1982). In addition, the *E. coli rplE* gene encoding Rpl5 is an essential gene as the knockout mutant is lethal (Korepanov et al. 2007). Rpl5 has been shown to play a crucial role in the formation of the central protuberance containing 5S rRNA and proteins L5, L16, L18, L25, L27, L31, L33, and L35 during assembly of the large ribosomal subunit in the bacterial cell (Korepanov et al. 2012). These observations indicate the importance of this ribosomal protein for ribosome formation and functioning. The existence of protein variants of Rpl5 in Arabidopsis generated by editing at a position that is phylogenetically conserved constitutes a puzzling result. However, the binding of Rpl5 to the 5S rRNA is mediated through one of the loops of this small RNA (Wang et al. 2013). It has been shown in vitro that Rpl5 from Arabidopsis is able to bind specifically to the potato spindle tuber viroid RNA that possesses a similar loop as the 5S rRNA (Eiras et al. 2011). One can then speculate that, like Rps12, Rpl5 might have taken on other functions in the mitochondrion than solely its involvement in the translation apparatus. Resolving the function of this protein polymorphism will require finding out whether both protein forms, histidine/unedited and tyrosine/edited, are incorporated in the mitochondrial ribosome.

In this study, we observed that most proteins found in the Arabidopsis organelles are translated from edited transcripts suggesting some control either cotranslationally or posttranslationally that excludes the presence of unedited proteins. In that respect, our results support the *raison d'être* of editing as a corrective mechanism that allows production of functional proteins. Nevertheless, we have also uncovered one instance of a protein polymorphism generated by editing in an essential protein, Rpl5. Whether this polymorphism has an impact on the function(s) of Rpl5 remains to be investigated.

PTMs of the plastid- and mitochondrial-encoded proteomes

Plant proteins undergo many PTMs to regulate protein function and stability (Friso and van Wijk 2015; Grabsztunowicz et al. 2017). Here we reported on 4 of the most important and studied PTMs, i.e. N-terminal protein acetylation, lysine ϵ-amine (side-chain) acetylation, phosphorylation (on S,T,Y), and ubiquitination, based on the large-scale reprocessing and searching of MS/MS data. Importantly, except for 1 observation for ubiquitination (plastid Rps7-K13), we did not identify any ubiquitination for plastid- or mitochondrial-encoded proteins (but we identified ∼1,000 ubiquitinated nuclear-encoded proteins). This is in line with our reassessment (van Wijk, Leppert, Sun, and Deutsch 2023) of a recent paper that claimed extensive polyubiquitination in chloroplasts (Sun et al. 2022). The lack of ubiquitination of plastid-encoded proteins is consistent with the lack of known E1, E2, and E3 enzymes involved in ubiquitin activation and ligation. We also did not observe any NTA of mitochondrial-encoded proteins consistent with a previous report (Huang et al. 2009) and the lack of known acetyl aminotransferases in mitochondria (Pozoga et al. 2022). In

contrast, 17 of the plastid-encoded proteins were *N*-acetylated at amino acid positions 1 and 2, consistent with the presence of multiple acetyl amino transferases (Pozoga et al. 2022) and previous results. However, the functional relevance of these NTA is not yet demonstrated but is likely to impact protein stability (e.g. through an N-degron pathway; Bouchnak and van Wijk 2019; Aguilar Lucero et al. 2021) or protein–protein interactions. The lack of observed NTA for the mitochondrial-encoded proteins, despite searching such a large set of MS/MS data (~257 million), further supports the notion that this PTM does not occur for mitochondrial-encoded proteins, and it also indicates a very low FDR in our data analysis workflow for this PTM.

Lysine ϵ-amine acetylation is driven by the presence of acetyl molecules as part of fatty acid metabolism and likely contributes to fine-tuning the activity of central metabolic enzymes, respiration, and photosynthetic activity, which likely impact plant acclimation capacity (Koskela et al. 2018; Balparda et al. 2022; Fussl et al. 2022). Here we identified extensive lysine acetylation of 16 plastid-encoded proteins (66 sites) and 2 mitochondrial proteins (Atp1 and Nad1; 4 sites). MS/MS analysis of leaf proteomes from Arabidopsis and other plant species (e.g. *Glycine max* [soybean] and rice) suggests that thousands of plant proteins undergo lysine ϵ-amine acetylation (Bienvenut et al. 2020; Li et al. 2021). Our data show that plastid-encoded proteins are also targets of protein acetylases. It remains to be determined what the biological impact is on these organellar proteins and which GNATs are responsible for these PTMs.

Phosphorylation has been observed for both plastid and mitochondrial localized proteins, in particular for nuclear-encoded proteins imported into these organelles (van Wijk et al. 2014; Willems et al. 2019; Zhang et al. 2021), and the thylakoid membrane system is well known to be regulated by phosphorylation (Rantala et al. 2020; Longoni and Goldschmidt-Clermont 2021). The single phosphorylated mitochondrial protein (Rps3 [pS112]) is important for mitochondrial biogenesis as loss of Rps3 splicing results in strong phenotypes (Sakamoto et al. 1996), but a role of phosphorylation of Rps3 has yet to be determined. The 12 plastid-encoded phosphorylated proteins are part of Photosystem II, Photosystem I, the ATP–synthase complex, the 30S ribosome, RbcL, and Tic214. Thylakoid protein (de) phosphorylation is governed by 2 state transition protein kinases (serine/threonine kinase 7,8 [STN7,8]) and 2 protein phosphatases (Protein Phosphatase 1 [PPH1/TAP38] and PSII core phosphatase [PBCP]), whereas plastid casein kinase (pCKII) is involved in the broader network of plastid protein phosphorylation (Rantala et al. 2020; Longoni and Goldschmidt-Clermont 2021). RbcL (pT333) and Rps7 (pS93) phosphorylation sites were also detected and shown to be partially dependent on kinase STN7 (Schonberg et al. 2017). Consistently, whereas we observed multiple phosphosites in RbcL, pT333 was by far the most frequently observed (68 PSMs). Of the previously reported N-terminal threonine p-sites on the plastid-encoded PSII core proteins, D1, D2, CP43, and PsbH upon exposure to high light (Rantala et al. 2020; Longoni and Goldschmidt-Clermont 2021), we did identify 2 closely spaced phosphorylated threonines on PsbH (T3 and T5; Supplemental Data Set 1). We also identified phosphorylation on 4 of the plastid-encoded subunits of the plastid ATP–synthase complex (Cfaα,β and Cfo-I,III) of which the 2 Cf1 subunits were also previously found to be phosphorylated (Sattari Vayghan et al. 2022).

Conclusion

This study is an important milestone in defining the plastid- and mitochondrial-encoded proteins with correct protein sequences, RNA editing state, and coverage of several physiological PTMs in Arabidopsis Col-0. The survey of organelle editing at a proteome-wide level shows that editing is nearly always required for stable protein accumulation, even for sites where RNA editing is well below 100%. These findings support the maintenance of this complex phenomenon in plants solely for the purpose of correcting deleterious mutations at the RNA level. This information should be incorporated in the next Arabidopsis genome build and annotation, and it will simplify MS-based proteome analysis of these organelle-encoded proteins. This study also demonstrates the benefits of reanalyzing publicly available MS/MS raw data sets at a large scale and shows the importance of careful verification of the underlying MS data for low-frequency events.

Materials and methods

The Arabidopsis PeptideAtlas build 2

The analysis in this study was based upon the ensemble results of the 2022-06 build of the Arabidopsis PeptideAtlas, as described in detail in van Wijk, Leppert, et al. (2023). Briefly, the raw data files from 115 data sets were downloaded from ProteomeXchange data repositories (Deutsch, Bandeira, et al. 2023) and reprocessed with the Trans-Proteomic Pipeline (Deutsch, Mendoza, et al. 2023) using parameters appropriate to each data set. This resulted in the identification of over 70 million spectra, nearly 0.6 million peptides, and 18,267 proteins (out of 27,559) at the highest confidence level.

Creating the protein search space and search strategy, allowing for partial editing

The protein search space included all protein sequences in Araport11, TAIR10, UniProtKB, Refseq, and others, as described in van Wijk, Leppert, et al. (2023). We also included the consensus set of 79 plastid-encoded and 33 mitochondrial-encodednonredundant protein sequences (Tables 1 and 2) to address mistakes and redundancies in plastid and mitochondrial ATGC and ATMG identifiers in Araport11 (same as in TAIR10; as explained further in Results). We also included variants of protein sequences for those plastid- and mitochondrial-encoded proteins that are predicted to be affected by RNA editing. For the mitochondrial-encoded

proteins, we included 420 editing sites in 29 mitochondrial-encoded proteins and 2 ORFs, most of which are described in Sloan et al. (2018) whereas we included edited sequences for 17 plastid-encoded proteins that included 31 amino acid changes and generation of 1 start methionine. These organellar-encoded sequences included unedited sequences and completely edited sequences. If predicted editing sites were sufficiently close together to appear in a single peptide (i.e. within ~30 amino acids of each other), we also included additional sequences that contained all possible permutations of nonedits and edits, to be able to detect instances of partial editing in 1 proteoform. This resulted in the addition of 10,368 sequences for plastid- and mitochondrial-encoded variants to the search database. These nonsynonymous editing sites were carefully assembled based on information from publications, public sequence databases, and discussions with experts in plant organelle RNA editing, in particular TAIR (Araport11 and TAIR10), UniProtKB, NCBI-Refgen, and *A. thaliana* Col-0 ecotype mitochondrial data set from Sloan et al. (2018), as well as plant RNA editing databases PREPACT3.0 (http://www.prepact.de/prepact-main.php), REDIdb RNA editing database (http://srv00.recas.ba.infn.it/redidb/), and the Plant editosome database (https://ngdc.cncb.ac.cn/ped/; Lenz et al. 2018; Lo Giudice et al. 2018; Li et al. 2019).

Manual verification of editing sites detected at low frequency

To ensure that editing sites observed in the proteome at low frequency (in their unedited or edited form) were correctly identified, matched MS/MS spectra were manually evaluated in PeptideAtlas build 2 to verify for the presence of y- and b-ions that specifically identified residue positions that could be affected by editing. For this editing site analysis, we also restricted possible variable mass modifications to methionine oxidation (but included semitryptic peptides) to reduce possible false-positive editing sites. The final reported results in this publication are made after these evaluations; hence, the spectra that were identified by the software but did not pass manual validation are still available in PeptideAtlas.

Accession numbers

Protein identification numbers are provided in the tables and supplemental tables.

Acknowledgments

We thank the plant organelle research community, in particular Ian Small, Josh Hazelwood, Etienne Meyer, and Ralph Bock, for information and feedback on organellar genes and editing sites.

Author contributions

T.L. and Z.S. carried out the MS searches and PeptideAtlas data loading, supervised by E.W.D., and assembled the search results. M.L. contributed to data analysis. L.M. and Z.S. developed the PeptideAtlas web interface. Q.S. helped assemble the protein search space. E.W.D., K.J.v.W., and S.B. developed the project, evaluated outcomes, and wrote the paper.

Supplemental data

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

Supplemental Figure S1. Primary protein sequence for NdhB/Ndh2 and conflicting reports on its N-terminus.

Supplemental Figure S2. Primary protein sequence for PsbC shows identical N-terminal sequences across PeptideAtlas, Tair10, Araport 11, UniProt, and NP_051055.

Supplemental Figure S3. Protein sequences of the 16 plastid-encoded proteins that are generated from spliced RNA and experimental MS support of splice junctions.

Supplemental Figure S4. Protein sequences of the 9 mitochondrial-encoded proteins that are generated from spliced RNA and experimental MS support of splice junctions.

Supplemental Data Set 1. Consensus annotation, nonsynonymous editing sites, protein identification, and editing status of the plastid-encoded proteins based on Arabidopsis PeptideAtlas build 2.

Supplemental Data Set 2. Consensus annotation, nonsynonymous editing sites, protein identification, and editing status of mitochondrial-encoded proteins based on Arabidopsis PeptideAtlas build 2.

Supplemental Data Set 3. Verified protein editing status of 156 editing sites in 27 mitochondrial-encoded proteins based manual evaluation of the second release of the Arabidopsis PeptideAtlas.

Supplemental Data Set 4. Detection of 4 physiologically important PTMs, N-terminal protein α-amine acetylation, lysine ϵ-amine acetylation, phosphorylation, and ubiquitination.

Funding

This research was support by a grant from National Science Foundation-IOS 1922871 to K.J.v.W. and E.W.D. and a grant from National Science Foundation-MCB 2122032 to S.B.

Conflict of interest statement. None declared.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

Aguilar Lucero D, Cantoia A, Sanchez-Lopez C, Binolfi A, Mogk A, Ceccarelli EA, Rosano GL. Structural features of the plant N-recognin ClpS1 and sequence determinants in its targets that govern substrate selection. FEBS Lett. 2021:**595**(11):1525–1541. https://doi.org/10.1002/1873-3468.14081

Balparda M, Elsasser M, Badia MB, Giese J, Bovdilova A, Hudig M, Reinmuth L, Eirich J, Schwarzlander M, Finkemeier I, et al. Acetylation of conserved lysines fine-tunes mitochondrial malate

dehydrogenase activity in land plants. Plant J. 2022:**109**(1):92–111. https://doi.org/10.1111/tpj.15556

Belinky F, Rogozin IB, Koonin EV. Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions. Sci Rep. 2017:**7**(1):12422. https://doi.org/10.1038/s41598-017-12619-6

Bentolila S, Babina AM, Germain A, Hanson MR. Quantitative trait locus mapping identifies REME2, a PPR-DYW protein required for editing of specific C targets in Arabidopsis mitochondria. RNA Biol. 2013:**10**(9):1520–1525. https://doi.org/10.4161/rna.25297

Bentolila S, Oh J, Hanson MR, Bukowski R. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. PLoS Genet. 2013:**9**(6):e1003584. https://doi.org/10.1371/journal.pgen.1003584

Bienvenut WV, Brunje A, Boyer JB, Muhlenbeck JS, Bernal G, Lassowskat I, Dian C, Linster E, Dinh TV, Koskela MM, et al. Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. Mol Syst Biol. 2020:**16**(7):e9464. https://doi.org/10.15252/msb.20209464

Bouchnak I, van Wijk KJ. N-degron pathways in plastids. Trends Plant Sci. 2019:**24**(10):917–926. https://doi.org/10.1016/j.tplants.2019.06.013

Cao X, Slavoff SA. Non-AUG start codons: expanding and regulating the small and alternative ORFeome. Exp Cell Res. 2020:**391**(1):111973. https://doi.org/10.1016/j.yexcr.2020.111973

Carrie C, Weissenberger S, Soll J. Plant mitochondria contain the protein translocase subunits TatB and TatC. J Cell Sci. 2016:**129**(20):3935–3947. https://doi.org/10.1242/jcs.190975

Chateigner-Boutin AL, des Francs-Small CC, Delannoy E, Kahlau S, Tanz SK, de Longevialle AF, Fujii S, Small I. OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. Plant J. 2010:**65**(4):532–542. https://doi.org/10.1111/j.1365-313X.2010.04441.x

Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, de la Luz Gutierrez-Nava M, Cantero A, Delannoy E, Jimenez LF, Lurin C, Small I, et al. CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. Plant J. 2008:**56**(4):590–602. https://doi.org/10.1111/j.1365-313X.2008.03634.x

Chotewutmontri P, Barkan A. Dynamics of chloroplast translation during chloroplast differentiation in maize. PLoS Genet. 2016:**12**(7):e1006106. https://doi.org/10.1371/journal.pgen.1006106

Coetzee T, Herschlag D, Belfort M. *Escherichia coli* proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. Genes Dev. 1994:**8**(13):1575–1588. https://doi.org/10.1101/gad.8.13.1575

Davidson NO. Apolipoprotein B mRNA editing: a key controlling element targeting fats to proper tissue. Ann Med. 1993:**25**(6):539–543. https://doi.org/10.1080/07853890.1993.12088581

Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu YZ, Weigel D, Mackenzie SA. Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol. 2011:**9**(1):64. https://doi.org/10.1186/1741-7007-9-64

Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver JJ, Mendoza L, Kundu DJ, Wang S, Bandla C, Kamatchinathan S, et al. The ProteomeXchange consortium at 10 years: 2023 update. Nucleic Acids Res. 2023:**51**(D1):D1539–D1548. https://doi.org/10.1093/nar/gkac1040

Deutsch EW, Mendoza L, Shteynberg DD, Hoopmann MR, Sun Z, Eng JK, Moritz RL. Trans-proteomic pipeline: robust mass spectrometry-based proteomics data analysis suite. J Proteome Res. 2023:**22**(4):1245–1254. https://doi.org/10.1021/acs.jproteome.2c00748

Dinh TV, Bienvenut WV, Linster E, Feldman-Salit A, Jung VA, Meinnel T, Hell R, Giglione C, Wirtz M. Molecular identification and functional characterization of the first Nalpha-acetyltransferase in plastids by global acetylome profiling. Proteomics 2015:**15**(14):2426–2435. https://doi.org/10.1002/pmic.201500025

Eiras M, Nohales MA, Kitajima EW, Flores R, Daros JA. Ribosomal protein L5 and transcription factor IIIA from *Arabidopsis thaliana* bind in vitro specifically potato spindle tuber viroid RNA. Arch Virol. 2011:**156**(3):529–533. https://doi.org/10.1007/s00705-010-0867-x

Friso G, van Wijk KJ. Posttranslational protein modifications in plant metabolism. Plant Physiol. 2015:**169**(3):1469–1487. https://doi.org/10.1104/pp.15.01378

Fuchs P, Rugen N, Carrie C, Elsasser M, Finkemeier I, Giese J, Hildebrandt TM, Kuhn K, Maurino VG, Ruberti C, et al. Single organelle function and organization as estimated from Arabidopsis mitochondrial proteomics. Plant J. 2020:**101**(2):420–441. https://doi.org/10.1111/tpj.14534

Fussl M, Konig AC, Eirich J, Hartl M, Kleinknecht L, Bohne AV, Harzen A, Kramer K, Leister D, Nickelsen J, et al. Dynamic light- and acetate-dependent regulation of the proteome and lysine acetylome of *Chlamydomonas*. Plant J. 2022:**109**(1):261–277. https://doi.org/10.1111/tpj.15555

Germain A, Hanson MR, Bentolila S. High-throughput quantification of chloroplast RNA editing extent using multiplex RT-PCR mass spectrometry. Plant J. 2015:**83**(3):546–554. https://doi.org/10.1111/tpj.12892

Grabsztunowicz M, Koskela MM, Mulo P. Post-translational modifications in regulation of chloroplast function: recent advances. Front Plant Sci. 2017:**8**:240. https://doi.org/10.3389/fpls.2017.00240

Green BR. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 2011:**66**(1):34–44. https://doi.org/10.1111/j.1365-313X.2011.04541.x

Grubb LE, Derbyshire P, Dunning KE, Zipfel C, Menke FLH, Monaghan J. Large-scale identification of ubiquitination sites on membrane-associated proteins in *Arabidopsis thaliana* seedlings. Plant Physiol. 2021:**185**(4):1483–1488. https://doi.org/10.1093/plphys/kiab023

Higashi H, Kato Y, Fujita T, Iwasaki S, Nakamura M, Nishimura Y, Takenaka M, Shikanai T. The pentatricopeptide repeat protein PGR3 is required for the translation of petL and ndhG by binding their 5′ UTRs. Plant Cell Physiol. 2021:**62**(7):1146–1155. https://doi.org/10.1093/pcp/pcaa180

Huang S, Taylor NL, Whelan J, Millar AH. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol. 2009:**150**(3):1272–1285. https://doi.org/10.1104/pp.109.137885

Korepanov AP, Gongadze GM, Garber MB, Court DL, Bubunenko MG. Importance of the 5 S rRNA-binding ribosomal proteins for cell viability and translation in *Escherichia coli*. J Mol Biol. 2007:**366**(4):1199–1208. https://doi.org/10.1016/j.jmb.2006.11.097

Korepanov AP, Korobeinikova AV, Shestakov SA, Garber MB, Gongadze GM. Protein L5 is crucial for in vivo assembly of the bacterial 50S ribosomal subunit central protuberance. Nucleic Acids Res. 2012:**40**(18):9153–9159. https://doi.org/10.1093/nar/gks676

Koskela MM, Brunje A, Ivanauskaite A, Grabsztunowicz M, Lassowskat I, Neumann U, Dinh TV, Sindlinger J, Schwarzer D, Wirtz M, et al. Chloroplast acetyltransferase NSI is required for state transitions in *Arabidopsis thaliana*. Plant Cell 2018:**30**(8):1695–1709. https://doi.org/10.1105/tpc.18.00155

Kuroda H, Suzuki H, Kusumegi T, Hirose T, Yukawa Y, Sugiura M. Translation of psbC mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine-Dalgarno sequence in tobacco chloroplasts. Plant Cell Physiol. 2007:**48**(9):1374–1378. https://doi.org/10.1093/pcp/pcm097

Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinformatics 2018:**19**(1):255. https://doi.org/10.1186/s12859-018-2244-9

Li G, Zheng B, Zhao W, Ren T, Zhang X, Ning T, Liu P. Global analysis of lysine acetylation in soybean leaves. Sci Rep. 2021:**11**(1):17858. https://doi.org/10.1038/s41598-021-97338-9

Li M, Xia L, Zhang Y, Niu G, Li M, Wang P, Zhang Y, Sang J, Zou D, Hu S, et al. Plant editosome database: a curated database of RNA editosome in plants. Nucleic Acids Res. 2019:**47**(D1):D170–D174. https://doi.org/10.1093/nar/gky1026

Lo Giudice C, Pesole G, Picardi E. REDIdb 3.0: a comprehensive collection of RNA editing events in plant organellar genomes. Front Plant Sci. 2018:**9**:482. https://doi.org/10.3389/fpls.2018.00482

Longoni FP, Goldschmidt-Clermont M. Thylakoid protein phosphorylation in chloroplasts. Plant Cell Physiol. 2021:**62**(7):1094–1107. https://doi.org/10.1093/pcp/pcab043

Lu B, Hanson MR. A single homogeneous form of ATP6 protein accumulates in petunia mitochondria despite the presence of differentially edited atp6 transcripts. Plant Cell 1994:**6**(12):1955–1968. https://doi.org/10.1105/tpc.6.12.1955

Lu B, Hanson MR. Fully edited and partially edited nad9 transcripts differ in size and both are associated with polysomes in potato mitochondria. Nucleic Acids Res. 1996:**24**(7):1369–1374. https://doi.org/10.1093/nar/24.7.1369

Lu B, Wilson RK, Phreaner CG, Mulligan RM, Hanson MR. Protein polymorphism generated by differential RNA editing of a plant mitochondrial rps12 gene. Mol Cell Biol. 1996:**16**(4):1543–1549. https://doi.org/10.1128/MCB.16.4.1543

Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria—past, present and future. Plant J. 2021:**108**(4):912–959. https://doi.org/10.1111/tpj.15495

Phreaner CG, Williams MA, Mulligan RM. Incomplete editing of rps12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. Plant Cell 1996:**8**(1):107–117. https://doi.org/10.1105/tpc.8.1.107

Planchard N, Bertin P, Quadrado M, Dargel-Graffin C, Hatin I, Namy O, Mireau H. The translational landscape of *Arabidopsis* mitochondria. Nucleic Acids Res. 2018:**46**(12):6218–6228. https://doi.org/10.1093/nar/gky489

Plochinger M, Schwenkert S, von Sydow L, Schroder WP, Meurer J. Functional update of the auxiliary proteins PsbW, PsbY, HCF136, PsbN, TerC and ALB3 in maintenance and assembly of PSII. Front Plant Sci. 2016:**7**:423. https://doi.org/10.3389/fpls.2016.00423

Pozoga M, Armbruster L, Wirtz M. From nucleus to membrane: a subcellular map of the N-acetylation machinery in plants. Int J Mol Sci. 2022:**23**(22):14492. https://doi.org/10.3390/ijms232214492

Rantala M, Rantala S, Aro EM. Composition, phosphorylation and dynamic organization of photosynthetic protein complexes in plant thylakoid membrane. Photochem Photobiol Sci. 2020:**19**(5):604–619. https://doi.org/10.1039/d0pp00025f

Rao RS, Salvato F, Thal B, Eubel H, Thelen JJ, Moller IM. The proteome of higher plant mitochondria. Mitochondrion 2016:**33**:22–37. https://doi.org/10.1016/j.mito.2016.07.002

Rohl R, Nierhaus KH. Assembly map of the large subunit (50S) of *Escherichia coli* ribosomes. Proc Natl Acad Sci U S A. 1982:**79**(3):729–733. https://doi.org/10.1073/pnas.79.3.729

Rojas M, Ruwe H, Miranda RG, Zoschke R, Hase N, Schmitz-Linneweber C, Barkan A. Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. Nucleic Acids Res. 2018:**46**(19):10448–10459. https://doi.org/10.1093/nar/gky737

Rowland E, Kim J, Bhuiyan NH, van Wijk KJ. The Arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. Plant Physiol. 2015:**169**(3):1881–1896. https://doi.org/10.1104/pp.15.01214

Roy S, Ueda M, Kadowaki K, Tsutsumi N. Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus Arabidopsis and closely related species. Genes Genet Syst. 2010:**85**(5):319–326. https://doi.org/10.1266/ggs.85.319

Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB. Arabidopsis chloroplast quantitative editotype. FEBS Lett. 2013:**587**(9):1429–1433. https://doi.org/10.1016/j.febslet.2013.03.022

Sadhu L, Kumar K, Kumar S, Dass A, Pathak R, Bhardwaj A, Pandey P, Van Cuu N, Rawat BS, Reddy VS. Chloroplasts evolved an additional layer of translational regulation based on non-AUG start codons for proteins with different turnover rates. Sci Rep. 2023:**13**(1):896. https://doi.org/10.1038/s41598-022-27347-9

Sakamoto W, Kondo H, Murata M, Motoyoshi F. Altered mitochondrial gene expression in a maternal distorted leaf mutant of Arabidopsis induced by chloroplast mutator. Plant Cell 1996:**8**(8):1377–1390. https://doi.org/10.1105/tpc.8.8.1377

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 1999:**6**(5):283–290. https://doi.org/10.1093/dnares/6.5.283

Sattari Vayghan H, Nawrocki WJ, Schiphorst C, Tolleter D, Hu C, Douet V, Glauser G, Finazzi G, Croce R, Wientjes E, et al. Photosynthetic light harvesting and thylakoid organization in a CRISPR/Cas9 *Arabidopsis thaliana* LHCB1 knockout mutant. Front Plant Sci. 2022:**13**:833032. https://doi.org/10.3389/fpls.2022.833032

Schonberg A, Rodiger A, Mehwald W, Galonska J, Christ G, Helm S, Thieme D, Majovsky P, Hoehenwarter W, Baginsky S. Identification of STN7/STN8 kinase targets reveals connections between electron transport, metabolism and gene expression. Plant J. 2017:**90**(6):1176–1186. https://doi.org/10.1111/tpj.13536

Seeburg PH, Hartner J. Regulation of ion channel/neurotransmitter receptor function by RNA editing. Curr Opin Neurobiol. 2003:**13**(3):279–283. https://doi.org/10.1016/S0959-4388(03)00062-X

Shteynberg DD, Deutsch EW, Campbell DS, Hoopmann MR, Kusebauch U, Lee D, Mendoza L, Midha MK, Sun Z, Whetton AD, et al. PTMProphet: fast and accurate mass modification localization for the trans-proteomic pipeline. J Proteome Res. 2019:**18**(12):4262–4272. https://doi.org/10.1021/acs.jproteome.9b00205

Sloan DB, Wu Z, Sharbrough J. Correction of persistent errors in Arabidopsis reference mitochondrial genomes. Plant Cell 2018:**30**(3):525–527. https://doi.org/10.1105/tpc.18.00024

Small I, Melonek J, Bohne AV, Nickelsen J, Schmitz-Linneweber C. Plant organellar RNA maturation. Plant Cell 2023:**35**(6):1727–1751. https://doi.org/10.1093/plcell/koad049

Small ID, Schallenberg-Rudinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020:**101**(5):1040–1056. https://doi.org/10.1111/tpj.14578

Sun Y, Yao Z, Ye Y, Fang J, Chen H, Lyu Y, Broad W, Fournier M, Chen G, Hu Y, et al. Ubiquitin-based pathway acts inside chloroplasts to regulate photosynthesis. Sci Adv. 2022:**8**(46):eabq7352. https://doi.org/10.1126/sciadv.abq7352

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA editing in plants and its evolution. Annu Rev Genet. 2013:**47**(1):335–352. https://doi.org/10.1146/annurev-genet-111212-133519

Tsugeki R, Kochieva EZ, Fedoroff NV. A transposon insertion in the *Arabidopsis* SSR16 gene causes an embryo-defective lethal mutation. Plant J. 1996:**10**(3):479–489. https://doi.org/10.1046/j.1365-313X.1996.10030479.x

Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, Kadowaki K. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. Mol Biol Evol. 2008:**25**(8):1566–1575. https://doi.org/10.1093/molbev/msn102

Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet. 1997:**15**(1):57–61. https://doi.org/10.1038/ng0197-57

van Wijk KJ, Friso G, Walther D, Schulze WX. Meta-analysis of *Arabidopsis thaliana* phospho-proteomics data reveals compartmentalization of phosphorylation motifs. Plant Cell 2014:**26**(6):2367–2389. https://doi.org/10.1105/tpc.114.125815

van Wijk KJ, Leppert T, Sun Q, Boguraev SS, Sun Z, Mendoza L, Deutsch EW. The Arabidopsis PeptideAtlas: harnessing worldwide proteomics data to create a comprehensive community proteomics resource. Plant Cell 2021:**33**(11):3421–3453. https://doi.org/10.1093/plcell/koab211

van Wijk KJ, Leppert T, Sun Z, Deutsch E. Does the ubiquitination degradation pathway really reach inside of the chloroplast? A re-evaluation of mass spectrometry-based assignments of ubiquitination. J Proteome Res. 2023:**22**(6):2079–2091. https://doi.org/10.1021/acs.jproteome.3c00178

van Wijk KJ, Leppert T, Sun Z, Kearly A, Li M, Mendoza L, Guzchenko I, Debley E, Sauermann G, Routray P, et al. Detection of the Arabidopsis proteome and its post-translational modifications and the nature of the unobserved (dark) proteome in PeptideAtlas. J Proteome Res. 2023. in press.

Walton A, Stes E, Cybulski N, Van Bel M, Inigo S, Durand AN, Timmerman E, Heyman J, Pauwels L, De Veylder L, et al. It's time for some "site"-seeing: novel tools to monitor the ubiquitin landscape in *Arabidopsis thaliana*. Plant Cell 2016:**28**(1):6–16. https://doi.org/10.1105/tpc.15.00878

Wang L, Ciganda M, Williams N. Defining the RNA-protein interactions in the trypanosome preribosomal complex. Eukaryot Cell. 2013:**12**(4):559–566. https://doi.org/10.1128/EC.00004-13

Willems P, Horne A, Van Parys T, Goormachtig S, De Smet I, Botzki A, Van Breusegem F, Gevaert K. The plant PTM viewer, a central resource for exploring plant protein modifications. Plant J. 2019:**99**(4):752–762. https://doi.org/10.1111/tpj.14345

Willems P, Ndah E, Jonckheere V, Van Breusegem F, Van Damme P. To new beginnings: riboproteogenomics discovery of N-terminal proteoforms in *Arabidopsis thaliana*. Front Plant Sci. 2021:**12**:778804. https://doi.org/10.3389/fpls.2021.778804

Williams MA, Tallakson WA, Phreaner CG, Mulligan RM. Editing and translation of ribosomal protein S13 transcripts: unedited translation products are not detectable in maize mitochondria. Curr Genet. 1998:**34**(3):221–226. https://doi.org/10.1007/s002940050390

Zhang Y, Giese J, Kerbler SM, Siemiatkowska B, de Souza L P, Alpers J, Medeiros DB, Hincha DK, Daloso DM, Stitt M, et al. Two mitochondrial phosphatases, PP2c63 and Sal2, are required for posttranslational regulation of the TCA cycle in Arabidopsis. Mol Plant. 2021:**14**(7):1104–1118. https://doi.org/10.1016/j.molp.2021.03.023

Zoschke R, Bock R. Chloroplast translation: structural and functional organization, operational control, and regulation. Plant Cell 2018:**30**(4):745–770. https://doi.org/10.1105/tpc.18.00016

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One 2008:**3**(4):e1994. https://doi.org/10.1371/journal.pone.0001994