HAL
open science

N-terminal modifications, the associated processing machinery, and their evolution in plastid-containing organisms

Thierry Meinnel, Carmela Giglione

To cite this version:

Thierry Meinnel, Carmela Giglione. N-terminal modifications, the associated processing machinery, and their evolution in plastid-containing organisms. Journal of Experimental Botany, 2022, 73 (18), pp.6013-6033. 10.1093/jxb/erac290 . hal-03754997

HAL Id: hal-03754997

https://hal.science/hal-03754997v1

Submitted on 20 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Darwin Review

Amino terminal modifications, the associated processing machinery, and their evolution in plastid-containing organisms

Thierry Meinnel[iD,@,1,*] and Carmela Giglione[iD,@,1,*]

[iD] ORCID iD 0000-0001-5642-8637 (T. Meinnel); 0002-7475-1558 (C. Giglione)

[@] twitter @meinnel (T. Meinnel); @giglionelab (C. Giglione)

[1] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France

***Correspondence:** thierry.meinnel@i2bc.paris-saclay.fr (T. Meinnel) or carmela.giglione@i2bc.paris-saclay.fr (C. Giglione)

Running Title: Amino terminal modifications and their catalysts in Plantae and Chromalveolata

STATEMENT

This review discusses N-termini modifications and their associated processing machinery in organisms with plastids primarily gained by primary or secondary endosymbiosis.

ABSTRACT

The N-terminus is a frequent site of protein modifications. Referring primarily to knowledge gained from land plants, here we review the modifications that change protein amino terminal residues and provide updated information about the associated machinery, including that in Archaeplastida. These N-terminal (Nt) modifications include many proteolytic events as well as small group additions such as acylation or arginylation and oxidation. Compared with that of the mitochondrion, the plastid-dedicated N-modification landscape is far more complex. In parallel, we extend this review to plastid-containing Chromalveolata including Stramenopiles, Apicomplexa, and Rhizaria. We report a well conserved machinery, especially in the plastid. Consideration of the two most abundant proteins on Earth - RuBisCO and actin - reveals the complexity of N-terminal modification processes. The progressive gene transfer from the plastid to the nuclear genome during evolution is exemplified by the N-terminus modification machinery, which appears to be one of the latest to have been transferred to the nuclear genome together with crucial major photosynthetic landmarks. This is evidenced by the greater number of plastid genes in Paulinellidae and red algae, the most recent and fossil recipients of primary endosymbiosis.

Keywords: acetylation, aminopeptidase, arginylation, deformylation, formylation, methylation, myristoylation, oxidation, plastid, proteolysis

N-terminal protein modifications in plastid-containing organisms in brief

During protein synthesis, protein N-termini are the first to emerge from the ribosome tunnel. When the nascent protein emerges, it immediately interacts with a welcoming committee that rapidly determines protein fate (Giglione *et al.*, 2015). Theoretically, all nascent chains undergo several modifications, which then allow them to undergo further modifications, interact with new partners, translocate or interact with specific compartments, or have longer or shorter half-lives. Unlike other parts of the protein, the N-terminus and any co-translational processes can be conditionally exposed to the crowed, highly interactive, and reactive space of small and large molecules in the cytosol, matrix, stroma, or even endoplasmic reticulum (ER), where they are available for additional post-translational protein modifications. The remarkable reactivity of the N-terminal amino group makes the N-terminal amino acid a substrate of choice for many enzymes capable of adding new groups, many of which cancel the positive charge. These enzymes often specifically recognize the N-terminal residue side chain and may sometimes also sense the protein sequence downstream.

Plastids have an endosymbiotic origin. At least two independent primary endosymbiosis events occurred during evolution, one ancestral ~1.6 billion years ago from a β-cyanobacterium leading to the Archaeplastida phylum with red alga the closest fossil, and another more recent event from an α-cyanobacterium leading to Paulinellidae ~100 million years ago. Except Paulinellidae, most Chromalveolatae have acquired their plastid more recently secondarily via phagocytosis of a red alga into another non-photosynthetic eukaryote. Dinoflagellates have a tertiary origin (Yoon *et al.*, 2005). Many protein modifications occurring in plastids as we currently understand them therefore arise from early endosymbiotic bacteria, which were progressively enslaved by a primitive eukaryote with a mitochondrion. A few complete and fully annotated plastid genomes for Chromalveolata have been released over the last few decades. These have revealed plastid genomes of various sizes

but a remarkable convergence between protein-encoding genes and evolutionary stage (**Box 1**).

Here we review N-terminal protein modifications in plastid-containing organisms. Plastid-containing organisms span all phyla in Bikonta (see **Figure 1**). Such organisms have gained their plastid in different ways and at different times during evolution, the earliest ancestor most likely soon after the last eukaryotic common ancestor ("LECA") gained the mitochondrion through early endosymbiosis. Plastid enslavement and the many interactions between the plastid and nuclear genomes are part of a complex and long-established process that profoundly affected not only metabolism towards autotrophy in most recipient organisms but also the protein modification landscape. We illustrate here several such modifications and how the timescale of plastid enslavement allowed the retention or migration of key elements from one genome to another (**Box 1**).

We first focus on N-terminal protein modifications, particularly those studies performed in plastid-containing organisms. In a dedicated section, we introduce the general features of methods that led to the discovery of protein modifications at the proteomic scale. We go on to describe the complex machinery in plastid-containing organisms and finally describe the complexity and varieties of these processes using actins and RuBisCO as exemplars.

The large diversity of N-terminal modifications in plastid-containing organisms

Proteins synthesized in plastid-containing organisms arise from three distinct cytosolic, plastid, and mitochondrial genomes. These proteins may shuttle between compartments, with the secretory pathway a rich source. In the cytosol, ribosomes distribute into free and ER-associated fractions, and in the latter case early interactions between the nascent chains and the secretory machinery can drive the association. As a result, cytosolic ribosomes synthesize mostly cytosolic proteins, whereas ER-bound ribosomes synthesize proteins mainly

corresponding to the secretory and integral membranes but also a few cytosolic proteins (Reid and Nicchitta, 2015). The molecular and modification landscapes of these two populations may therefore be strongly affected.

Co-translational amino terminal processing

Protein synthesis in the cytosol leads to the incorporation of an N-terminal methionine supported by the sole AUG codon. However, ribosome profiling has frequently retrieved AUG start codons downstream or upstream of the annotated AUG (Li and Liu, 2020). As proteomic approaches rely on the genomic annotation of open reading frames, care must therefore be taken in their interpretation when dealing with N-termini. This is even truer for genomes relying on less data than model organisms such as the plant dicot *Arabidopsis thaliana*. For instance, while 19% of N-termini in *A. thaliana* are wrongly annotated, 31% of tomato N-termini are wrongly annotated; no corresponding data are yet available for Chromalveolata. In addition, it was established early on that there were inconsistencies in 5' genome annotations in Apicomplexa (Wakaguri *et al.*, 2009). Whilst usually encoded by the AUG codon, various non-AUG codons (e.g., CUG, ACG, GUG, AUU) may also initiate proteins, particularly in plants (up to 35% of translation starts in tomato) (Li and Liu, 2020). Nevertheless, such codons invariably lead to the incorporation of methionine (Met); indeed, the initiator Met-tRNA itself directs initiation, even though imperfect or wobble codon-anticodon pairing may occur (Na *et al.*, 2018).

Organelles use a bacterial-type translation machinery and, unlike the cytosolic mechanism, which relies on archaeal-type machinery, the initiator Met-tRNA must first be modified by a formyl group prior to protein synthesis. As a result, a pre-modified N-formylated methionine is incorporated early in translation. This unique protein modification is not usually present in mature proteins as it undergoes early cleavage (Giglione *et al.*, 2004). In bacteria, for instance, about 5% of proteins retain their N-terminus, most of which are membrane proteins

(Bienvenut *et al.*, 2015). This probably means that there is competition between co-translational N-terminal protein modifications and the early processing of such proteins by the secretory machinery including the signal recognition particle. While it remains unclear whether formyl is retained or removed in metazoan and fungal mitochondria, in plant mitochondria, the N-terminal formyl group is rarely displayed (Giglione and Meinnel, 2001). In humans, N-formyl must be retained on cytochrome *c* oxidase subunit COX1 in the mitochondrion for complex assembly (Hinttala *et al.*, 2015). A few mature formylated proteins might therefore be expected in the plastids and mitochondria of plastid-containing organisms (Giglione *et al.*, 2004). Unlike in plants, there are still no data on formylated proteins in Chromalveolata.

Though not removed as frequently as the formyl group, the N-terminal methionine is also generally cleaved off (Giglione *et al.*, 2004; Giglione and Meinnel, 2001; Millar *et al.*, 2019). This process of N-terminal methionine excision (NME) is conserved in all living organisms studied to date. About 60% of cytosolic and plastid proteins undergo NME (Giglione *et al.*, 2004), and it is an essential process in the cytosol of *A. thaliana* (Ross *et al.*, 2005) and the plastids of monocots and dicots (Dirk *et al.*, 2001; Giglione *et al.*, 2003; Moon *et al.*, 2008; Serero *et al.*, 2001). Both steps of plastid NME are also essential in Apicomplexa (Bushell *et al.*, 2017). The molecular mechanisms underpinning NME and its importance have been studied in both the cytosol and plastids of plants. In the cytosol, NME controls proteostasis by ensuring tightly regulated crosstalk between the cellular proteolytic machinery and glutathione homeostasis (Frottin *et al.*, 2009). Specifically, it was shown that retention of the N-terminal methionine, by blocking NME, increases the cellular proteolytic machinery by increasing the pool of proteins suitable for rapid degradation. Very interestingly, N-terminal-retained methionines were very often found to be unacetylated when NME was inhibited, suggesting that the unacetylated methionines could act as an N-degron. This is consistent with

a recent finding on the non-Ac-degron (see below and (Linster *et al.*, 2022)), and plastid NME has been suggested to protect a specific pool of proteins from rapid degradation with retention of the N-terminal methionine interpreted as a destabilizing signal (Adam *et al.*, 2011; Giglione *et al.*, 2003).

NME is also important for further modifications that fully rely on NME, such as N-terminal acetylation (Nt-acetylation) and N-myristoylation. Nt-acetylation and N-myristoylation describe the addition of short or long aliphatic acetyl or acyl (C2 and C14, respectively) groups to N-termini (Giglione and Meinnel, 2021, 2022) by specific enzymes belonging to the Gcn5-related N-acetyltransferase (GNAT) superfamily: N-acetyl-transferases (NATs) and N-myristoyl-transferases (NMTs). These acyl additions make the N-terminus hydrophobic and cancel the positive charge. Both modifications are co-translational, as they occur soon after the nascent chain emerges from the ribosome, probably to ensure complete modification of the N-terminus. As a result, Nt-acetylation is a protein modification leading to completion for the majority of substrates in the cytosol, as shown in quantitative analyses performed in plants (Huber *et al.*, 2020; Linster *et al.*, 2015). Nt-acetylation has also been described for the plastid, where it modifies both nascent proteins arising from translation of the plastid genome and proteins imported from the cytosol. In the latter case, the modification is present on the proteolytically-processed new N-terminus devoid of the transit peptide (Bienvenut *et al.*, 2011; Bienvenut *et al.*, 2012). Unlike in co-translation, there is only partial post-translational modification of nuclear-encoded imported proteins - 30% on average – although this might be as high as 100% in some cases. By contrast, it is remarkable that complete Nt-acetylation occurs on most plastid-encoded proteins such as photosystem 2 components (e.g., D1 or D2) or the large subunit of RuBisCO (RbcL); however, no data are yet available on whether this is a co-translational process, like in bacteria. Though not yet described in plastid-containing organisms, post-translational Nt-acetylation and N-myristoylation have been detected in the

cytosol of metazoans (Giglione and Meinnel, 2022; Lange *et al.*, 2014). One cannot exclude, therefore, that proteins from plastid-containing organisms undergoing proteolysis in the cytosol or other subcellular compartments might further undergo N-myristoylation and Nt-acetylation.

Nt-acetylation is frequent in both the cytosol and plastids of plants, and the modification has also been retrieved in Stramenopiles (diatoms) and Rhizaria (Huesgen *et al.*, 2013; Oberleitner *et al.*, 2022). By contrast, N-myristoylation occurs in about 1-2% of proteins. N-myristoylation has been studied in detail in *A. thaliana*, in which nearly all the modified proteome has now been characterized (Castrec *et al.*, 2018; Majeran *et al.*, 2018). N-myristoylation studies using a clickable precursor have also been performed in *Toxoplasma gondii*, where 65 targets were identified and many others predicted (Alonso *et al.*, 2019; Broncel *et al.*, 2020). The substrate specificity of *Plasmodium falciparum* NMT was studied using an array of octapeptides, and it was shown to display close similarities with *A. thaliana* and human NMT, while the yeast isoform appeared to feature slightly more restricted selectivity (Pierre *et al.*, 2007; Traverso *et al.*, 2013a). N-myristoylation usually allows its protein targets to anchor the membrane towards the cytosol. The transmembrane microneme protein MIC7 was shown to display N-myristoylation, with the myristoylated N-terminus facing the lumen (Broncel *et al.*, 2020). This suggests that the modified N-terminus might flip from the outer to the inner side of the membrane, possibly during co-translational insertion of the transmembrane domains.

N-myristoylation has been studied in detail in *A. thaliana*. Membrane specificity not only depends on N-terminal myristoylation but also vicinal features including palmitoylation of cysteines next to the modified N-terminal glycine (Boisson *et al.*, 2003; Majeran *et al.*, 2018; Pierre *et al.*, 2007; Traverso *et al.*, 2013b). In plants, the essential nature of N-myristoylation likely depends on a set of major kinases and small G-proteins such as ADP-ribosyl factor

(ARF) at various crucial development stages, including fertilization and germination (Bayer *et al.*, 2009; Pierre *et al.*, 2007; Renna *et al.*, 2013). Evolutionary investigations have also revealed a small number of fully conserved NMT targets in all organisms including ARFs (Meinnel *et al.*, 2020). Overall, the features of Nt-acetylation and N-myristoylation appear to be similar in metazoans and plastid-containing organisms, indicating the early acquisition of the system after the LECA and before primary endosymbiosis of the β-cyanobacterium leading to the primary plastid ~1.6 billion years ago.

Intracellular proteolysis leads to new N-termini

Since most mitochondrial genes and many plastid genes have migrated to the nuclear genome during evolution, proteins transit from the cytosol to organelles by taking advantage of targeting peptides. This phenomenon is well studied and described in plants (Christian *et al.*, 2020). This transit usually leads to proteolytic cleavages by dedicated machineries, similar to proteins transiting via the secretory pathway. In fact, recent data indicate that it is very likely that the secretory pathway is used as a means for proteins to cross the three to four membrane layers that separate the plastid from the cytosol in Chromalveolata, in which plastids originate from (i) secondary endosymbiosis (i.e., most phyla including Cercozoa like *Bigelowiella natans* (Hopkins *et al.*, 2012) but not ciliates and foraminifera, **Figure 1**) or (ii) primary endosymbiosis (only Paulinellidae in the phylum). It appears that the ER signaling peptide must be fused to a transit peptide for protein import into complex plastids like those in the diatom *Thalassiosira pseudonana* (Huesgen *et al.*, 2013). The situation in Apicomplexa is most similar to that in non-photosynthetic plastid- (so-called apicoplast) targeting sequences (Tonkin *et al.*, 2008). The apicoplast also has a secondary origin and is separated from the cytosol by four membranes. Like diatoms, Apicomplexa use a target peptide formed from the fusion of an ER- and plastid-targeting sequence. In addition to the plastid, these organisms contain advanced secretory compartments including rhoptries, micronemes, or dense granules,

which are specialized for parasite-host interactions (Sheiner and Soldati-Favre, 2008). Consequently, complex proteolytic events occur in these unique compartments, leading to new downstream cleavages of secreted proteins. These involve dedicated proteases including aminopeptidase N and aspartyl protease ASP5, which we will not discuss further (Coffey *et al.*, 2018; Dogga *et al.*, 2017). Additional modifications including acetylation have also been described in this pathway (Nyonda *et al.*, 2022).

Methods of investigation and the discovery of protein modifications

Protein modifications and modified proteomes

A range of approaches have now been used over many years to discover and annotate modified proteomes (Meinnel and Giglione, 2008), including methods based on: (i) precursors tagged with radioactive or reactive species including clickable moieties (Kallemeijn *et al.*, 2021; Suazo *et al.*, 2021) or featuring unique chemical bond sensitivity (e.g., acyl biotin exchange for S-acylation); (ii) enrichments of modified peptides using affinity approaches such as Strong Cation eXchange columns (SCX; e.g., phosphorylation) or monoclonal antibodies (e.g., acetylation); (iii) *in vitro* challenges of the catalyst with sorted or random alanine-scanning libraries when the enzyme modifier is known; (iv) affinity-based approaches such as proximity assays to visualize and identify protein modification substrates (e.g., methylation (Poulard *et al.*, 2012)); (v) site-directed variants with substitutions abolishing target modification site reactivity or mimicking the modified target; and (vi) genome-wide mass spectrometry approaches with dedicated surveys for specific modifications (e.g., ubiquitination with GGK tags (Parker *et al.*, 2019)) (Aebersold and Mann, 2016). Complementary bioinformatics approaches are integral to progress in this field, particularly when the modification is frequent and complex, for instance in the case of palmitoylation (Blanc *et al.*, 2015). Target catalogs are listed in several repositories and can be searched. Annotations in protein databanks remain incomplete, for instance in the

reference protein resource UniProtKB (Protein KnowledgeBase), but focused efforts to integrate these data are currently underway (Zahn-Zabal *et al.*, 2020).

In vitro and *in vivo* approaches have frequently been combined with structural and bioinformatics analyses to further complete the picture. *In vitro* approaches are often used upstream to help rationalize the data obtained from *in vivo* samples, which may sometimes be difficult to interpret due to sample preparation artefacts such as oxidation of thiol-containing residues or chemical acetylation of lysines (Baeza *et al.*, 2015). Furthermore, the sample may be complex, sometimes producing billions of overlapping signals. The interpretation of those approaches that use reactive precursors can also be ambiguous, as the precursor may be metabolized into other compounds used by other catalysts on different targets or lead to weak signals that do not represent the real stoichiometry or biology. Lysine acetylation, for instance, is most often associated with low stoichiometry (see Fig. 7 in (Lindahl *et al.*, 2019) and (Hansen *et al.*, 2019; O'Leary *et al.*, 2020; Prus *et al.*, 2019)). All data arising from these diverse approaches therefore provide qualitative information about modified subproteomes. However, the quantitative assessment of modified proteomes remains in its infancy, as stoichiometry, competition with other modifications, tissue and environmental heterogeneity, or isoform heterogeneity require characterization and are usually poorly documented.

The special case of N-terminal modifications and their discovery

The discovery of position-specific modifications becomes even more challenging. These modifications need to be identified on only one peptide and not randomly on any peptide whatever its position. This restricts the identification search space to an average of one out of 20-40 tryptic peptides. For instance, a given protein has only one N-terminus and an average of 32 lysine acceptors available for trypsin cleavage. In addition, classic proteomic label-free approaches cannot be used for protein quantification and stable isotopes need to be employed. To retrieve proteome-wide information, N-terminal peptides must be enriched. Specific

properties of the N-terminal peptide are therefore exploited for this purpose. One major property relies on NH_2 terminal group reactivity with aldehyde derivatives. The amino terminal group is often acetylated in eukaryotes or can react with dedicated modifiers. Although the pK is much lower than that of lysines and would theoretically allow the labeling of only one group if the reaction is at the proper pH, most protocols lead to the modification of both amino groups. This was certainly the case in early approaches such as SILProNAQ (Bienvenut *et al.*, 2017; Bienvenut *et al.*, 2012) or COFRADIC, which are readily applied to plants (Tsiatsiani *et al.*, 2014; Tsiatsiani *et al.*, 2013). Improved workflows have now been described for plants (Venne *et al.*, 2015). Here we only summarize the main features and refer interested readers to an excellent recent review on the topic in plants (Perrar *et al.*, 2019).

The chemical labeling reaction takes place before proteins are cleaved by proteases like trypsin, which cleaves at lysines and arginines. As a result, this blocks lysine cleavage by trypsin and produces longer peptides. Some peptides cannot be analyzed in this way because their length is incompatible with mass spectrometry. In these cases, a good alternative is the use of other proteases such as endoproteinase GluC, which cleaves after glutamates (Zhang *et al.*, 2018). Terminal Amine Isotopic Labelling of Substrates (TAILS) coupled with dimethyl labeling is one commonly used approach (Rowland *et al.*, 2015; Zhang *et al.*, 2015). The HUNTER workflow (Weng *et al.*, 2019), which also relies on chemical dimethylation, was recently used to examine *Paulinella chromatophora* processing in the chromatophore (**Figure 2**). Internal peptides react with a hydrophobic chain (undecanal, a C11 aliphatic aldehyde), and the N-terminal peptides can be enriched by reverse phase chromatography. Stable isotope labeling is used when quantification is required, for instance deuterated acetic acid for acetylation quantification. Tandem mass tag labeling has been implemented in the TAILS workflow (Zhang *et al.*, 2018). Another alternative is to use click chemistry if precursors can be incorporated in the growth medium. So far, however, unlike for other organisms such as

fungi or metazoans, few large-scale approaches have been described for plastid-containing organisms, probably because they display almost complete autotrophy. For instance, N-myristoylation has been reported for several proteins using azido or alkyne derivatives (Boyle *et al.*, 2016; Lee *et al.*, 2021). Catalytic parameters obtained *in vitro* - including k_{cat}/K_m values – of N-terminal peptides help to categorize substrates and the reaction as exemplified by the systematic analysis of *A. thaliana* N-myristoylation (Castrec *et al.*, 2018).

The processing machinery of plastid-containing organisms

In this section, we discuss current knowledge about the catalysts that specifically mediate N-terminal protein processing and their associated specificity in plastid-containing organisms. The main modification events are summarized in **Figure 3**. We use representative members of each phylum (**Figure 1**) to provide examples of the corresponding machinery (see also **Tables 1** and **2**).

The green plant *A. thaliana* lies in the Archaeplastida phylum. In the Chromalveolata phylum, which includes Stramenopiles, Alveolata, and Rhizaria (SAR), we use the oleaginous alga *Nannochloropsis gaditana* - which features lipid-production capacity up to six-fold higher than those of other algae (Radakovits *et al.*, 2012) - as a model of Stramenopiles. We also use *Paulinella* species to exemplify the Rhizaria phylum and *Toxoplasma gondii* - a non-photosynthetic organism - as a model of Alveolata (**Figure 1**). All organisms were chosen because of the superior quality of their genome annotation and assembly. We do not discuss Euglenozoa (see further information in the "Concluding remarks").

Proteolytic events promoting new protein N-termini

Plants display a plethora of peptidases impacting protein N-termini leading to new N-terminal protein modifications (**Table 1**). These include the dedicated endoproteases that cleave off the signal/targeting signals in the ER, mitochondria, and plastid, namely signal peptide peptidase

(SIP/SIPL), the mitochondrial-processing peptidase (MPP), and the stromal processing peptidase (SPP). The substrate specificity appears rather large, and the activity is completed by further aminopeptidases (APP) that trim the N-terminus. The nature of these APPs is not completely understood, but some of them (e.g., MPA1, LAP2/3, and DAP2) have been found in the *A. thaliana* plastid, where they may be involved in a multistep processes leading to progressive trimming of new N-terminal moieties (Teixeira *et al.*, 2017) (**Table 1**). It is interesting that a number of these proteases show significant similarity to the *PepP* bacterial aminopeptidase gene product, which also shares homology with MetAPs - the conserved component of the NME machinery (see below) - which promotes specific N-terminal processing in bacteria likely in combination with MetAP.

In Chromalveolata, the transit peptide of plastid-targeted proteins is bipartite to ensure ER and then plastid retargeting, and it appears to be cleaved by homologs of SPP. Diatoms and Apicomplexa including *T. gondii* (van Dooren *et al.*, 2002) encode an unambiguous SPP homolog (**Table 1**). No such SPP homolog has been detected in the rhizarian Paulinellidae, which might explain why only partial cleavage of the long targeting peptide has been retrieved (Oberleitner *et al.*, 2022). SIP and MPP homologs have been identified in diatoms and Apicomplexa (**Table 1**)

Of the APPs, MetAPs have more restricted specificity and are involved in NME. Dedicated forms appear in the organelles and cytosol in plants (Giglione *et al.*, 2000b). Cytosolic isoforms have an N-terminal extension featuring polybasic or zinc-binding domains, both of which are believed to be involved in ribosome targeting. In the plastid, NME is initiated by deformylation, as MetAPs cannot cleave the formyl group. Peptide deformylases (PDFs) unmask the N-formyl methionine for methionine removal by MetAPs (Giglione *et al.*, 2000b). Though not formally true proteases, PDF's action is closely related to that of an aminopeptidase, and PDFs display the same HEXXH motif in the catalytic center as many

proteases (Giglione *et al.*, 2000a). Both MetAPs and PDF homologs can be retrieved in Stramenopiles, Apicomplexa, and Rhizaria. In Paulinellidae, PDFs and MetAPs are encoded within the chromatophore (**Box 1**). Finally, pyroglutamyl peptidases can cleave pyroglutamate in the cytosol.

Addition reactions at the N-terminus

The mechanisms leading to the addition of a chemical group to the N-terminus involve a large range of modifications.

Formylation

Though incorporated early in protein translation, we will consider N-formylation as the very first modification revealed by any protein. This modification relies on Met-tRNA^{fMet} formyltransferase (FMT), an enzyme that specifically modifies both mitochondrial and plastid initiator tRNAs using 10-formyltetrahydrofolate as the methyl donor. The enzyme is conserved in diatoms and encoded in the nuclear genome, where it features a transit peptide. The situation is different in other Chromalveolata such as Apicomplexa, as (i) in *Toxoplasma* FMT appears to be the second part of a two-domain protein, and (ii) as yet there is no evidence of an FMT homolog in *Plasmodium*, whose genome is significantly smaller than other Apicomplexa. In Paulinellidae, the gene is encoded in the chromatophore - the plastid with a large genome - as a remnant of early endosymbiosis (**Box 1**). In *B. natans*, a Cercozoa with plastids of secondary origin - no homolog has been found. In *A. thaliana*, FMT is essential, as gene inactivation leads to an albino phenotype (pigment-defective *pde194*; (Bryant *et al.*, 2011)). Although it may also be essential in Chromalveolata, organisms in the phylum lacking an FMT homolog might be less dependent on the efficient organelle translation initiation expected from formylation due to the smaller plastid genome, as seen in Apicomplexa or parasitism (see also **Box 1**).

Acetylation

N-terminal acetylation (Nt-acetylation) is the most common N-terminus modification. The modification is ensured by a complex machinery which is similar in most eukaryotes, and the catalytic subunits can be identified in most plastid-containing organisms including Chromalveolata, Archaeplastida, and Euglenozoa (**Table 2**). The many different NATs and their associated cell localization, substrate specificity, and activity on cellular proteins are conserved in eukaryotes and are detailed in **Figure 4A**. It is remarkable, however, that some important ancillary subunits, including HYPK of the NatA complex (except in diatoms) or NAA35 of the NatC complex in Apicomplexa, are missing or could not be identified in Chromalveolata. The same is also true for Paulinellidae; while an NAA40 homolog could not be retrieved, there was acetylation of its specific histone substrates. Remarkably, Paulinellidae also have NAA80 activity, as actin derivatives display N-terminal acidic acid (i.e., Glu or Asp) acetylation, a modification which appears also to occur in *Toxoplasma* but not in *Plasmodium* or plants (**Table 2**). NAA80 homologs have not been identified in plastid-containing organisms, but one cannot exclude that this activity arises from another type of NAT; for instance, human NAA10 devoid of the NAA15 subunit displayed such activity (Van Damme *et al.*, 2011). Finally, it is worth mentioning that recent data from large-scale N-terminal analyses have indicated that N-terminal acetylation also occurs in the Apicomplexa secretory pathway. Though the origin of AcetylCoA still needs to be established, it likely originates from a NAT, but the specific enzyme responsible has yet to be identified (Nyonda *et al.*, 2022). In *A. thaliana*, NatA, NatB, and NatE appear to be essential complexes as demonstrated in knockout experiments of the various subunits of each complex (Armbruster *et al.*, 2020; Ferrandez-Ayela *et al.*, 2013; Huber *et al.*, 2020; Linster *et al.*, 2015). NatC inactivation is associated with photosynthetic defects (Pesaresi *et al.*, 2003), while NAA60 is associated with high stress responses (Linster *et al.*, 2020).

The Plantae phylum displays a well conserved acetylation machinery (NatG) dedicated to the plastid (Bienvenut *et al.*, 2020). There are two NatG subfamilies, NAA90 and NAA70 (Giglione and Meinnel, 2021). Unlike NAA70, NAA90 homologs are found in red algae and the corresponding gene is usually encoded in the plastid (except in some Glaucophyta such as *Cyanophora paradoxa*). This is in keeping with only partial nuclear relocalization of the plastid genome in red algae (**Box 1**). In addition, this indicates that NAA90 is the ancestral NatG; as it can only be identified in cyanobacteria, it has likely evolved from the RimIJL family seen in all bacteria to mediate an important but as yet unknown acetylation required by the photosynthetic apparatus. Accordingly, two NatG homologs can be found in the plastid genome of Paulinellidae, which underwent very recent endosymbiosis from an α-cyanobacterium (Meinnel, 2022). This classification is related to the RbcL type phylogeny, which strongly relies on N-terminus differences. Representative members of the four subtypes (A-D) of type 1 RbcL N-termini and the known associated modifications are aligned in **Table 3**. NAA70 homologs – but not NAA90 – have been detected in diatoms and other Cercozoa. However, it is currently unknown which forms control plastid-resident Nt-acetylation or where the modification takes place.

Myristoylation

N-myristoyltransferase (NMT) is the only enzyme known to catalyze N-myristoylation in eukaryotes. NMT uses cytosolic myristoyl CoA to promote co-translational N-myristoylation (Giglione and Meinnel, 2022). In plastid-containing organisms, fatty acid synthesis is performed in the plastid. Myristate probably needs to travel from the plastid through the ER to be eventually available as MyrCoA in the cytosol for NMT-catalyzed N-myristoylation. In parasites such as Apicomplexa, depending on their amebic or parasitic form, the origin of myristate could differ, as most organisms can pump in fatty acids from the medium including the host's cytosol or blood. This is unlike autotrophic organisms such as diatoms, which only

rely on plastid synthesis. NMTs have been identified in Apicomplexa but not yet in Paulinellidae, though a homolog was identified in another Cercozoan (*Bigelowiella natans*, **Table 2**). N-myristoylation is a typical modification whose appearance during evolution is linked to early eukaryogenesis, most likely by fusion of two NAA catalytic subunits (Meinnel *et al.*, 2020). N-myristoylation, surprisingly, does not occur in the plastid despite being the site of fatty acid synthesis in plastid-containing organisms, while other organisms only rely on food uptake to fuel their cytosolic needs (Giglione and Meinnel, 2022).

Arginylation

N-terminal arginylation is an essential component of the N-degron protein degradation pathways (Varshavsky, 2019). Arginine is added on amino-terminal acidic residues such as oxidized Cys, Asp, or Glu - either native or arising from oxidation of Asn or Gln (see below). This arginylation guides targeted proteins to the proteasome complex, where the protein is degraded. Arginyl-tRNA-protein transferases (ATE) promote the transfer of arginyl from the Arg-tRNA to the protein. The ATE knockout mutant, initially identified as the delayed leaf senescence *dls1* mutant in *A. thaliana* (Yoshida *et al.*, 2002), actually displays a broad range of phenotypes including enhanced seed dormancy and improved hypoxia tolerance (see phenotypes and associated references in Figure 2 in (Dissmeyer, 2019)). We retrieved ATE homologs in Chromalveolata including diatoms and some Cercozoa (*Plasmodiophora brassicae*) but none yet in Paulinellidae and Apicomplexa (**Table 2**).

Methylation

Methylation is a rare modification catalyzed by protein N-methyltransferases (NTMs), with cytosolic activity in both plants and diatoms. Protein methylation has been reported in the plastids of plants for a couple of plastid encoded-ribosomal proteins but also for the small subunit of RuBisCO after cleavage of its transit peptide. An NTM (RBCMT) was proposed to mediate RbcS monomethylation (Houtz *et al.*, 2008; Ying *et al.*, 1999), but to date no catalyst

has been identified for the ribosomal activity (**Table 2**). Note that RBCMT should not be confused with another NTM (LSMT) that appears to be mostly involved in Lys14 trimethylation of RbcL. Homologs of cytosolic NTM exist in diatoms and Apicomplexa, and a homolog of the plastid RBCMT exists in diatoms. Modifications by NTM appear to affect N-terminal small residues such as Ala, Ser, or Pro followed by Pro-Lys (**Table 2**). Such a specific pattern restricts the modified proteome to only few targets.

Oxidative events occurring at the N-terminal residue

The amino terminal residue may undergo specific oxidation as a result of enzyme-catalyzed or chemical processes (Davies, 2005). Proteome-wide protocols usually start with protein oxidation leading to Met and Cys oxidation. They are therefore poorly informative about the oxidative state of the proteins. Dedicated approaches for protein oxidation have been established for plants, for instance during oxidative stress (Jacques *et al*., 2015). These have revealed that methionines, especially those of the plastid proteome, are a hotspot of oxidative stress. Despite trends to retain protein integrity with repair systems, these processes are closely associated with protein homeostasis - so-called proteostasis - and they contribute to protein half-life by driving them specifically to degradation.

<u>Methionine sulfoxide reduction</u>

Methionine is extremely sensitive to oxidation of its sulfur group by reactive oxygen species into sulfenic, sulfinic, sulfonic, or sulfoxide forms (Davies, 2005). Oxidation may lead to two stereoisomers. Though methionines are often retrieved proteolytically from the N-terminus, as noted above, a significant number are kept and often protected by acetylation, for instance by NatB, NatC, NatE, or NatH.

All cells are equipped with a dedicated salvage repair enzyme machinery called peptide-methionine-S-oxide reductase (MRS), which includes two catalysts (MRSA and MRSB), each

specific for either the *S* or *R* stereoisomer, respectively. Plastid-containing organisms also contain this machinery (**Table 2**), though we could not identify it in Paulinellidae. Plants possess an extremely diverse machinery localized in all cell compartments including highly oxidative organelles like plastids and mitochondria. However, the specific mitochondrial machinery remains to be identified in *A. thaliana*. Mitochondrial forms have been identified in other plants (Tarrago *et al.*, 2009). Overall, the machinery contributes to the general reduction of this residue (Rouhier *et al.*, 2006) and it is conserved in Archaeplastida (Tarrago *et al.*, 2009). Whether N-terminal methionines are more sensitive to oxidation than internal methionines, possibly because they are more exposed to the outer, more oxidative environment, is unknown. Internal methionines are rather buried and protected because of the amphipathic but more hydrophobic character of this linear side chain.

Oxidation of cysteines

Like mammals, plants display an active enzyme machinery that contributes to the oxidation of cysteine-sensitive N-terminal residues. This includes Nt-Cys cysteine oxidases, known as plant cysteine oxidases (PCOs) in plants (Graciet *et al.*, 2010; Vicente *et al.*, 2019; Weits *et al.*, 2014; White *et al.*, 2017). Cysteine oxidases have not been retrieved in Chromalveolata, and whether they use other devices to achieve this aim is unknown. In plants, this oxidation machinery was shown to be directly related to protein degradation, as it is sensed by N-degron degradation pathways in the cytosol (Gibbs *et al.*, 2016; Vicente *et al.*, 2019; Weits *et al.*, 2014).

The Nt-Cys oxidation and degradation system serves as sensor of oxygen concentration (Dissmeyer, 2019; Hammarlund *et al.*, 2020). Under normal oxygen concentrations, Nt-Cys oxidation leads to Nt arginylation by ATE and recognition by ubiquitin ligases such as PRT6 in plants, which target the protein for proteasome degradation. Under hypoxia - for instance triggered by flooding or waterlogging in plants - the Nt-Cys is no longer oxidized and the

protein is not degraded, which promotes a hypoxic response. This response involves ERF-VII transcription factors but also other targets like the polycomb repressive complex 2 subunit VERNALIZATION 2 or LITTLE ZIPPER 2 (Gibbs *et al.*, 2018; Holdsworth and Gibbs, 2020; Weits *et al.*, 2019).

Other N-terminal modifications

Gln and Asn deamidations

Nt-Asn (NTAN) and Nt-Gln (NTAQ) deamidases catalyze the conversion of Asn and Gln residues into Asp and Glu (Graciet *et al.*, 2010; Vicente *et al.*, 2019; Weits *et al.*, 2014; White *et al.*, 2017). Nt amidases have been identified in Stramenopiles and Rhizaria but not yet in Paulinellidae and Apicomplexa. In the plant cytosol, Nt deamidation is sensed by N-degron degradation pathways (Gibbs *et al.*, 2016; Vicente *et al.*, 2019; Weits *et al.*, 2014).

Pyroglutamate cyclization

Pyroglutamyl formation corresponds to the spontaneous cyclization of glutamate or glutamine residues in peptides and proteins. This cyclization is usually located at the N-terminus and leads to -18 and -17 Da mass shifts (Liu *et al.*, 2011; Schilling *et al.*, 2007; Schilling *et al.*, 2008). Cyclases in some plants, such as in the latex of *Carica papaya*, accelerate the reaction. There also exist aminopeptidases that can cleave off the N-terminal pyroglutamate residue (**Table 1**). Homologs of these peptidases can be retrieved in diatoms and Apicomplexa (**Table 1**).

The remarkable and unique examples of N-terminal modifications of actins and RuBisCO, the most abundant proteins on Earth

In this last section, we describe two major targets - the two most abundant proteins on Earth actins and RuBisCO (Ellis, 1979; Raven, 2013) - and their associated protein machinery as exemplars of N-terminal modifications. RuBisCO is a plastid-resident major complex

associated with CO_2 uptake, scavenging, and autotrophy in plastid-containing organisms, and actin is a major cytoskeletal component. The respective N-terminal sequences and associated modifications in plastid-containing organisms are depicted in **Tables 3 and 4** for the RbcL catalytic component of RuBisCO and actin isoforms, respectively. The data show that these abundant proteins have independently developed specific modification schemes that mostly mimic each other: trimming up to residue 3, while starts on residues 1 or 2 are the rule for most mature proteins relying on NME. This unique processing involves not only Nt-acetylation and NME but also processing of the second residue and unusual Nt-acetylation events (**Figure 4B**; **Tables 3** and **4**). The overall mechanism has yet to be deciphered, although recent data have described, for instance, the dedicated Nt-acetylation machinery in mammals (Drazic *et al.*, 2018; Drazic *et al.*, 2022). **Figure 4B** proposes the mechanism allowing the cytosol to achieve this remarkable unique scheme for actin. In mammals, this involves an N-acylaminopeptidase (ANAP) removing the NatB-acetylated N-terminal methionine. Mammalian ANAP consists of a purified and characterized 77 kDa monomer (Sheff and Rubenstein, 1992), but the associated gene remains elusive. Whilst plants and diatoms do not apparently display this unique actin modification pattern, Apicomplexa and Cercozoa do (**Table 4**). The associated machinery has yet to be clearly discovered but could mimic the mammalian one.

Similarly, a dedicated RbcL processing machinery occurs in plastid-containing organisms, type 1B RuBisCO (**Table 4**). The machinery leading to this unique processing has yet to be determined, but some hypotheses have been proposed (Houtz *et al.*, 2008). Note that dinoflagellates (Alveolata; **Figure 1**) feature a type II RbcL encoded in the nuclear genome and whose active form corresponds to a homodimer of RbcL without RbcS (Tabita *et al.*, 2008). Finally, it is remarkable that the small RbcS subunit of RuBisCO is also encoded in the plastids of many unicellular plastid-containing organisms but has relocated to the nucleus of

land plants (**Box 1**). In this latter situation, the transit peptide of its precursor is cleaved off and further monomethylated (Houtz *et al.*, 2008). In any case, the active type 1B RuBisCO quaternary structure is an octamer of the large (RbcL) and the small (RbcS) subunits (LS8).

Overall, it is remarkable that the most abundant proteins have developed unique N-terminal protein modification mechanisms. Whether this is due to their abundance and related stability is unknown.

Concluding remarks and future perspectives

Here we discussed the remarkable conservation but diversity of the modification machinery and proteomes of plastid-containing organisms. We are fully aware that we have not discussed Euglenozoa, a photosynthetic plastid-containing organism in the Excavata phylum (**Figure 1**). This was deliberate, not least due to the generally poor knowledge about these organisms, especially with respect to N-terminal protein modifications. We identified NAA10, NAA20, NAA30, and NAA50 homologs in *Euglena gracilis* and NMT in *Euglena longa*. A recent genome-wide integrated analysis of *E. gracilis* will be helpful for more accurate analyses of the associated N-terminal protein modification machinery in Euglenozoa (Ebenezer *et al.*, 2019).

N-terminal protein modifications generally serve various crucial purposes, primarily dedicated to cellular localization. They also act as unambiguous degradation signals in plants (Dissmeyer, 2019; Gibbs *et al.*, 2016). How and if this is also true in the plastid is unknown and is currently under investigation (Bouchnak and van Wijk, 2019). In the cytosol, very recent data have shown that not only oxidation but also Nt-acetylation contribute to protein stability and are crucial for stress adaptation; this appears to be a role of the small ancillary subunit HYPK of the NatA complex, which was not retrieved in Chromalveolata. Plant HYPK plays a crucial role in plant adaptation to non-marine environments (Gong *et al.*, 2022;

Linster *et al.*, 2022). This responsive role involving Nt-acetylation is also true for plant adaptation to pathogenic attack (Xu *et al.*, 2015). The unique adaptation of land plants to aerial life involves exposure to drought, salt stress, waterlogging, and flooding that may differentially affect distinct areas of different organs (roots *vs* shoots, for instance). Chromalveolata are usually unicellular marine, fresh water, or parasitic organisms that possibly do not suffer similar events or may adapt differently through protection from their siliceous shell, Finally, N-terminal ubiquitination is an N-terminal protein modification that has been described in mammals as a modification that contributes to shorten protein half-life (Meinnel *et al.*, 2005). So far, however, no such modification has been described in plastid-containing organisms, including plants and diatoms.

ACKNOWLEDGEMENTS

The team's research is funded by the French Agence Nationale de la Recherche (ANR-13-BSV6-0004, ANR-17-CAPS-0001-01, ANR-20-CE44-0013, ANR-20 CE92-0040), Fondation ARC (ARCPJA32020060002137), and has benefited from a French State grant (Saclay Plant Sciences, reference n° ANR-17-EUR-0007, EUR SPS-GSR) managed by the French National Research Agency under an Investments for the Future program (reference n° ANR-11-IDEX-0003-02) and the COST Action CA20113 ProteoCure.

AUTHOR CONTRIBUTION

TM and CG wrote the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

586

REFERENCES

Adam Z, Frottin F, Espagne C, Meinnel T, Giglione C. 2011. Interplay between N-terminal methionine excision and FtsH protease is essential for normal chloroplast development and function in Arabidopsis. The Plant Cell **23**, 3745-3760.

Aebersold R, Mann M. 2016. Mass-spectrometric exploration of proteome structure and function. Nature **537**, 347-355.

Aksnes H, Drazic A, Marie M, Arnesen T. 2016. First things first: vital protein marks by N-terminal acetyltransferases. Trends in Biochemical Sciences **41**, 746-760.

Alonso AM, Turowski VR, Ruiz DM, Orelo BD, Moresco JJ, Yates JR, 3rd, Corvi MM. 2019. Exploring protein myristoylation in Toxoplasma gondii. Experimental Parasitology **203**, 8-18.

Armbruster L, Linster E, Boyer JB, Brünje A, Eirich J, Stephan I, Bienvenut WV, Weidenhausen J, Meinnel T, Hell R, Sinning I, Finkemeier I, Giglione C, Wirtz M. 2020. NAA50 is an enzymatically active N(α)-acetyltransferase that is crucial for development and regulation of stress responses. Plant Physiology **183**, 1502-1516.

Aryal UK, McBride Z, Chen D, Xie J, Szymanski DB. 2017. Analysis of protein complexes in Arabidopsis leaves using size exclusion chromatography and label-free protein correlation profiling. Journal of Proteomics **166**, 8-18.

Aryal UK, Xiong Y, McBride Z, Kihara D, Xie J, Hall MC, Szymanski DB. 2014. A proteomic strategy for global analysis of plant protein complexes. The Plant Cell **26**, 3867-3882.

Baeza J, Smallegan MJ, Denu JM. 2015. Site-specific reactivity of nonenzymatic lysine acetylation. ACS Chemical Biology **10**, 122-128.

Baudisch B, Klösgen RB. 2012. Dual targeting of a processing peptidase into both endosymbiotic organelles mediated by a transport signal of unusual architecture. Molecular Plant **5**, 494-503.

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W. 2009. Paternal control of embryonic patterning in Arabidopsis thaliana. Science **323**, 1485-1488.

Bhuiyan NH, Rowland E, Friso G, Ponnala L, Michel EJS, van Wijk KJ. 2020. Autocatalytic Processing and Substrate Specificity of Arabidopsis Chloroplast Glutamyl Peptidase. Plant Physiology **184**, 110-129.

Bienvenut WV, Brünje A, Boyer J-B, Mühlenbeck JS, Bernal G, Lassowskat I, Dian C, Linster E, Dinh TV, Koskela MM, Jung V, Seidel J, Schyrba LK, Ivanauskaite A, Eirich J, Hell R, Schwarzer D, Mulo P, Wirtz M, Meinnel T, Giglione C, Finkemeier I. 2020. Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. Molecular Systems Biology **16**, e9464.

Bienvenut WV, Espagne C, Martinez A, Majeran W, Valot B, Zivy M, Vallon O, Adam Z, Meinnel T, Giglione C. 2011. Dynamics of post-translational modifications and protein stability in the stroma of Chlamydomonas reinhardtii chloroplasts. Proteomics **11**, 1734-1750.

Bienvenut WV, Giglione C, Meinnel T. 2015. Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition. Proteomics **15**, 2503-2518.

Bienvenut WV, Giglione C, Meinnel T. 2017. SILProNAQ: a convenient approach for proteome-wide analysis of protein N-termini and N-terminal acetylation quantitation. Methods in Molecular Biology **1574**, 17-34.

Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinnel T, Giglione C. 2012. Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. Molecular & Cellular Proteomics **11**, M111 015131.

Blanc M, David F, Abrami L, Migliozzi D, Armand F, Burgi J, van der Goot FG. 2015. SwissPalm: protein palmitoylation database. F1000Research **4**, 261.

Boisson B, Giglione C, Meinnel T. 2003. Unexpected protein families including cell defense components feature in the N-myristoylome of a higher eukaryote. The Journal of Biological Chemistry **278**, 43418-43429.

Bouchnak I, van Wijk KJ. 2019. N-degron pathways in plastids. Trends in Plant Science **24**, 917-926.

Boyle PC, Schwizer S, Hind SR, Kraus CM, De la Torre Diaz S, He B, Martin GB. 2016. Detecting N-myristoylation and S-acylation of host and pathogen proteins in plants using click chemistry. Plant Methods **12**, 38.

Broncel M, Dominicus C, Vigetti L, Nofal SD, Bartlett EJ, Touquet B, Hunt A, Wallbank BA, Federico S, Matthews S, Young JC, Tate EW, Tardieux I, Treeck M. 2020. Profiling of myristoylation in Toxoplasma gondii reveals an N-myristoylated protein important for host cell penetration. eLife **9**, e57861.

Bryant N, Lloyd J, Sweeney C, Myouga F, Meinke D. 2011. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. Plant Physiology **155**, 1678-1689.

Bushell E, Gomes AR, Sanderson T, Anar B, Girling G, Herd C, Metcalf T, Modrzynska K, Schwach F, Martin RE, Mather MW, McFadden GI, Parts L, Rutledge GG, Vaidya AB, Wengelnik K, Rayner JC, Billker O. 2017. Functional profiling of a Plasmodium genome Reveals an abundance of essential genes. Cell **170**, 260-272.e268.

Carrie C, Venne AS, Zahedi RP, Soll J. 2015. Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in Arabidopsis thaliana. Journal of Experimental Botany **66**, 2691-2708.

Castrec B, Dian C, Ciccone S, Ebert CL, Bienvenut WV, Le Caer J-P, Steyaert J-M, Giglione C, Meinnel T. 2018. Structural and genomic decoding of human and plant myristoylomes reveals a definitive recognition pattern. Nature Chemical Biology **14**, 671-679.

Christian RW, Hewitt SL, Nelson G, Roalson EH, Dhingra A. 2020. Plastid transit peptides-where do they come from and where do they all belong? Multi-genome and pan-genomic assessment of chloroplast transit peptide evolution. PeerJ **8**, e9772.

Coffey MJ, Dagley LF, Seizova S, Kapp EA, Infusini G, Roos DS, Boddey JA, Webb AI, Tonkin CJ. 2018. Aspartyl protease 5 matures dense granule proteins that reside at the host-parasite interface in *Toxoplasma gondii*. mBio **9**, e01796-01718.

Cook RK, Sheff DR, Rubenstein PA. 1991. Unusual metabolism of the yeast actin amino terminus. The Journal of Biological Chemistry **266**, 16825-16833.

Davies MJ. 2005. The oxidative environment and protein damage. Biochimica et Biophysica Acta **1703**, 93-109.

Dinh TV, Bienvenut WV, Linster E, Feldman-Salit A, Jung VA, Meinnel T, Hell R, Giglione C, Wirtz M. 2015. Molecular identification and functional characterization of the first Nalpha-acetyltransferase in plastids by global acetylome profiling. Proteomics **15**, 2426-2435.

Dirk LM, Williams MA, Houtz RL. 2001. Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast- targeted enzymes in *Arabidopsis*. Plant Physiology **127**, 97-107.

Dissmeyer N. 2019. Conditional protein function via N-degron pathway-mediated proteostasis in stress physiology. Annual Review of Plant Biology **70**, 83-117.

Dogga SK, Mukherjee B, Jacot D, Kockmann T, Molino L, Hammoudi PM, Hartkoorn RC, Hehl AB, Soldati-Favre D. 2017. A druggable secretory protein maturase of Toxoplasma essential for invasion and egress. eLife **6**, e27480.

Drazic A, Aksnes H, Marie M, Boczkowska M, Varland S, Timmerman E, Foyn H, Glomnes N, Rebowski G, Impens F, Gevaert K, Dominguez R, Arnesen T. 2018. NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility. Proceedings of the National Academy of Sciences **115**, 4399-4404.

Drazic A, Timmerman E, Kajan U, Marie M, Varland S, Impens F, Gevaert K, Arnesen T. 2022. The final maturation state of β-actin involves N-terminal acetylation by NAA80, not N-terminal arginylation by ATE1. Journal of Molecular Biology **434**, 167397.

Ebenezer TE, Zoltner M, Burrell A, Nenarokova A, Novák Vanclová AMG, Prasad B, Soukal P, Santana-Molina C, O'Neill E, Nankissoor NN, Vadakedath N, Daiker V, Obado S, Silva-Pereira S, Jackson AP, Devos DP, Lukeš J, Lebert M, Vaughan S, Hampl V, Carrington M, Ginger ML, Dacks JB, Kelly S, Field MC. 2019. Transcriptome, proteome and draft genome of Euglena gracilis. BMC Biology **17**, 11.

Ellis RJ. 1979. The most abundant protein in the world. Trends in Biochemical Sciences **4**, 241-244.

Ferrandez-Ayela A, Micol-Ponce R, Sanchez-Garcia AB, Alonso-Peral MM, Micol JL, Ponce MR. 2013. Mutation of an Arabidopsis NatB N-alpha-terminal acetylation complex component causes pleiotropic developmental defects. PloS One **8**, e80697.

Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, Rolland N. 2010. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Molecular & Cellular Proteomics **9**, 1063-1084.

Flecken M, Wang H, Popilka L, Hartl FU, Bracher A, Hayer-Hartl M. 2020. Dual functions of a Rubisco activase in metabolic repair and recruitment to carboxysomes. Cell **183**, 457-473.

Forsberg J, Ström J, Kieselbach T, Larsson H, Alexciev K, Engström Å, Åkerlund H-E. 2005. Protease activities in the chloroplast capable of cleaving an LHCII N-terminal peptide. Physiologia Plantarum **123**, 21-29.

Frottin F, Espagne C, Traverso JA, Mauve C, Valot B, Lelarge-Trouverie C, Zivy M, Noctor G, Meinnel T, Giglione C. 2009. Cotranslational proteolysis dominates glutathione homeostasis to support proper growth and development. The Plant Cell **21**, 3296-3314.

Gibbs DJ, Bailey M, Tedds HM, Holdsworth MJ. 2016. From start to finish: amino-terminal protein modifications as degradation signals in plants. The New Phytologist **211**, 1188-1194.

Gibbs DJ, Tedds HM, Labandera AM, Bailey M, White MD, Hartman S, Sprigg C, Mogg SL, Osborne R, Dambire C, Boeckx T, Paling Z, Voesenek L, Flashman E, Holdsworth MJ. 2018. Oxygen-dependent proteolysis regulates the stability of angiosperm polycomb repressive complex 2 subunit VERNALIZATION 2. Nature Communications **9**, 5438.

Giglione C, Boularot A, Meinnel T. 2004. Protein N-terminal methionine excision. Cellular and Molecular Life Sciences **61**, 1455-1474.

Giglione C, Fieulaine S, Meinnel T. 2015. N-terminal protein modifications: bringing back into play the ribosome. Biochimie **114**, 134-146.

Giglione C, Meinnel T. 2001. Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. Trends in Plant Science **6**, 566-572.

Giglione C, Meinnel T. 2021. Evolution-driven versatility of N terminal acetylation in photoautotrophs. Trends in Plant Science **26**, 375-391.

Giglione C, Meinnel T. 2022. Mapping the myristoylome through a complete understanding of protein myristoylation biochemistry. Progress in Lipid Research **85**, 101139.

Giglione C, Pierre M, Meinnel T. 2000a. Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. Molecular Microbiology **36**, 1197-1205.

Giglione C, Serero A, Pierre M, Boisson B, Meinnel T. 2000b. Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. The EMBO Journal **19**, 5916-5929.

Giglione C, Vallon O, Meinnel T. 2003. Control of protein life-span by N-terminal methionine excision. The EMBO Journal **22**, 13-23.

Gong X, Huang Y, Liang Y, Yuan Y, Liu Y, Han T, Li S, Gao H, Lv B, Huang X, Linster E, Wang Y, Wirtz M, Wang Y. 2022. OsHYPK-mediated protein N-terminal acetylation coordinates plant development and abiotic stress responses in rice. Molecular Plant **15**, 740-754.

Graciet E, Mesiti F, Wellmer F. 2010. Structure and evolutionary conservation of the plant N-end rule pathway. The Plant Journal **61**, 741-751.

Graciet E, Walter F, Ó'Maoiléidigh DS, Pollmann S, Meyerowitz EM, Varshavsky A, Wellmer F. 2009. The N-end rule pathway controls multiple functions during Arabidopsis shoot and leaf development. Proceedings of the National Academy of Sciences of the United States of America **106**, 13618-13623.

Hammarlund EU, Flashman E, Mohlin S, Licausi F. 2020. Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. Science **370**.

Hansen BK, Gupta R, Baldus L, Lyon D, Narita T, Lammers M, Choudhary C, Weinert BT. 2019. Analysis of human acetylation stoichiometry defines mechanistic constraints on protein regulation. Nature Communications **10**, 1055.

Hernández-Cuevas NA, Jhingan GD, Petropolis D, Vargas M, Guillen N. 2019. Acetylation is the most abundant actin modification in Entamoeba histolytica and modifications of actin's amino-terminal domain change cytoskeleton activities. Cellular Microbiology **21**, e12983.

Hinttala R, Sasarman F, Nishimura T, Antonicka H, Brunel-Guitton C, Schwartzentruber J, Fahiminiya S, Majewski J, Faubert D, Ostergaard E, Smeitink JA, Shoubridge EA. 2015. An N-terminal formyl methionine on COX 1 is required for the assembly of cytochrome c oxidase. Human Molecular Genetics **24**, 4103-4113.

Holdsworth MJ, Gibbs DJ. 2020. Comparative Biology of Oxygen Sensing in Plants and Animals. Current Biology **30**, R362-r369.

Hopkins JF, Spencer DF, Laboissiere S, Neilson JAD, Eveleigh RJM, Durnford DG, Gray MW, Archibald JM. 2012. Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga *Bigelowiella natans*. Genome Biology and Evolution **4**, 1391-1406.

Houtz RL, Magnani R, Nayak NR, Dirk LM. 2008. Co- and post-translational modifications in Rubisco: unanswered questions. Journal of Experimental Botany **59**, 1635-1645.

Houtz RL, Poneleit L, Jones SB, Royer M, Stults JT. 1992. Posttranslational Modifications in the Amino- Terminal Region of the Large Subunit of Ribulose- 1,5-Bisphosphate Carboxylase/Oxygenase from Several Plant Species. Plant Physiology **98**, 1170-1174.

Huang S, Nelson CJ, Li L, Taylor NL, Ströher E, Peteriet J, Millar AH. 2015. INTERMEDIATE CLEAVAGE PEPTIDASE55 Modifies Enzyme Amino Termini and Alters Protein Stability in Arabidopsis Mitochondria. Plant Physiology **168**, 415-427.

Huang S, Taylor NL, Whelan J, Millar AH. 2009. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiology **150**, 1272-1285.

Huber M, Bienvenut WV, Linster E, Stephan I, Armbruster L, Sticht C, Layer DC, Lapouge K, Meinnel T, Sinning I, Giglione C, Hell R, Wirtz M. 2020. NatB-mediated N-terminal acetylation affects growth and abiotic stress responses. Plant Physiology **182**, 792-806.

Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR. 2013. Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. PloS One **8**, e74483.

Inoue K, Baldwin AJ, Shipman RL, Matsui K, Theg SM, Ohme-Takagi M. 2005. Complete maturation of the plastid protein translocation channel requires a type I signal peptidase. The Journal of Cell Biology **171**, 425-430.

Jacques S, Ghesquière B, De Bock PJ, Demol H, Wahni K, Willems P, Messens J, Van Breusegem F, Gevaert K. 2015. Protein methionine sulfoxide dynamics in Arabidopsis thaliana under oxidative stress. Molecular & Cellular Proteomics **14**, 1217-1229.

Kallemeijn WW, Lanyon-Hogg T, Panyain N, Goya Grocin A, Ciepla P, Morales-Sanfrutos J, Tate EW. 2021. Proteome-wide analysis of protein lipidation using chemical probes: in-gel fluorescence visualization, identification and quantification of N-myristoylation, N- and S-acylation, O-cholesterylation, S-farnesylation and S-geranylgeranylation. Nature Protocols **16**, 5083-5122.

Kmiec B, Glaser E. 2012. A novel mitochondrial and chloroplast peptidasome, PreP. Physiologia Plantarum **145**, 180-186.

Lange PF, Huesgen PF, Nguyen K, Overall CM. 2014. Annotating N termini for the human proteome project: N termini and Nα-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. Journal of Proteome Research **13**, 2028-2044.

Lee ES, Park JH, Wi SD, Kang CH, Chi YH, Chae HB, Paeng SK, Ji MG, Kim WY, Kim MG, Yun DJ, Stacey G, Lee SY. 2021. Redox-dependent structural switch and CBF activation confer freezing tolerance in plants. Nature Plants.

Lensch M, Herrmann RG, Sokolenko A. 2001. Identification and characterization of SppA, a novel light-inducible chloroplast protease complex associated with thylakoid membranes. The Journal of Biological Chemistry **276**, 33645-33651.

Li YR, Liu MJ. 2020. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. Genome Research **30**, 1418-1433.

Lindahl AJ, Lawton AJ, Baeza J, Dowell JA, Denu JM. 2019. Site-specific lysine acetylation stoichiometry across subcellular compartments. Methods in Molecular Biology **1983**, 79-106.

Linster E, Forero Ruiz FL, Miklankova P, Ruppert T, Mueller J, Armbruster L, Gong X, Serino G, Mann M, Hell R, Wirtz M. 2022. Cotranslational N-degron masking by acetylation promotes proteome stability in plants. Nature Communications **13**, 810.

Linster E, Layer D, Bienvenut WV, Dinh TV, Weyer FA, Leemhuis W, Brünje A, Hoffrichter M, Miklankova P, Kopp J, Lapouge K, Sindlinger J, Schwarzer D, Meinnel T, Finkemeier I, Giglione C, Hell R, Sinning I, Wirtz M. 2020. The Arabidopsis Nα-acetyltransferase NAA60 locates to the plasma membrane and is vital for the high salt stress response. The New Phytologist **228**, 554-569.

Linster E, Stephan I, Bienvenut WV, Maple-Grodem J, Myklebust LM, Huber M, Reichelt M, Sticht C, Geir Moller S, Meinnel T, Arnesen T, Giglione C, Hell R, Wirtz M. 2015. Downregulation of N-terminal acetylation triggers ABA-mediated drought responses in Arabidopsis. Nature Communications **6**, 7640.

Liu YD, Goetze AM, Bass RB, Flynn GC. 2011. N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. The Journal of Biological Chemistry **286**, 11211-11217.

Lynch M, Field MC, Goodson HV, Malik HS, Pereira-Leal JB, Roos DS, Turkewitz AP, Sazer S. 2014. Evolutionary cell biology: two origins, one objective. Proceedings of the National Academy of Sciences of the United States of America **111**, 16990-16994.

Majeran W, Le Caer JP, Ponnala L, Meinnel T, Giglione C. 2018. Targeted profiling of A. thaliana sub-proteomes illuminates new co- and post-translationally N-terminal myristoylated proteins. The Plant Cell **30**, 543-562.

Meinnel T. 2022. Tracking N-terminal protein processing from the Golgi to the chromatophore of a rhizarian amoeba. Plant Physiology **in press, 10.1093/plphys/kiac173**.

Meinnel T, Dian C, Giglione C. 2020. Myristoylation, an ancient protein modification mirroring eukaryogenesis and evolution. Trends in Biochemical Sciences **45**, 619-632.

Meinnel T, Giglione C. 2008. Tools for analyzing and predicting N-terminal protein modifications. Proteomics **8**, 626-649.

Meinnel T, Peynot P, Giglione C. 2005. Processed N-termini of mature proteins in higher eukaryotes and their major contribution to dynamic proteomics. Biochimie **87**, 701-712.

Miklánková P, Linster E, Boyer JB, Weidenhausen J, Mueller J, Armbruster L, Lapouge K, De La Torre C, Bienvenut W, Sticht C, Mann M, Meinnel T, Sinning I, Giglione C, Hell R, Wirtz M. 2022. HYPK promotes the activity of the essential Nα-acetyltransferase A complex to determine proteostasis of nonAc-X^2/N-degron containing proteins. Science Advances **in press**.

Millar AH, Heazlewood JL, Giglione C, Holdsworth MJ, Bachmair A, Schulze WX. 2019. The scope, functions, and dynamics of posttranslational protein modifications. Annual Review of Plant Biology **70**, 119-151.

Moon S, Giglione C, Lee DY, An S, Jeong DH, Meinnel T, An G. 2008. Rice peptide deformylase PDF1B is crucial for development of chloroplasts. Plant and Cell Physiology **49**, 1536-1546.

Murphy AS, Hoogner KR, Peer WA, Taiz L. 2002. Identification, purification, and molecular cloning of N-1-naphthylphthalmic acid-binding plasma membrane-associated aminopeptidases from Arabidopsis. Plant Physiology **128**, 935-950.

Na CH, Barbhuiya MA, Kim MS, Verbruggen S, Eacker SM, Pletnikova O, Troncoso JC, Halushka MK, Menschaert G, Overall CM, Pandey A. 2018. Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Research **28**, 25-36.

Nakai A, Yamauchi Y, Sumi S, Tanaka K. 2012. Role of acylamino acid-releasing enzyme/oxidized protein hydrolase in sustaining homeostasis of the cytoplasmic antioxidative system. Planta **236**, 427-436.

Nyonda AM, Boyer J-B, Belmudes L, Krishnan A, Pino P, Couté Y, Brochet M, Meinnel T, Soldati-Favre D, Giglione C. 2022. N-Acetylation of secreted proteins is widespread in Apicomplexa and independent of acetyl-CoA transporter AT1. Journal of Cell Science **in press**.

O'Leary BM, Scafaro AP, Fenske R, Duncan O, Ströher E, Petereit J, Millar AH. 2020. Rubisco lysine acetylation occurs at very low stoichiometry in mature Arabidopsis leaves: implications for regulation of enzyme function. The Biochemical Journal **477**, 3885-3896.

Oberleitner L, Perrar A, Macorano L, Huesgen PF, Nowack ECM. 2022. A bipartite chromatophore transit peptide and N-terminal protein processing in the Paulinella chromatophore. Plant Physiology **189**, 152-164.

Park SY, Scranton MA, Stajich JE, Yee A, Walling LL. 2017. Chlorophyte aspartyl aminopeptidases: Ancient origins, expanded families, new locations, and secondary functions. PloS One **12**, e0185492.

Parker CE, Warren Hines MRE, Mocanu V, Greer SF, Borchers CH. 2019. Mass spectrometric determination of protein ubiquitination. Methods in Molecular Biology **1934**, 191-221.

Perrar A, Dissmeyer N, Huesgen PF. 2019. New beginnings and new ends: methods for large-scale characterization of protein termini and their use in plant biology. Journal of Experimental Botany **70**, 2021-2038.

Pesaresi P, Gardner NA, Masiero S, Dietzmann A, Eichacker L, Wickner R, Salamini F, Leister D. 2003. Cytoplasmic N-terminal protein acetylation is required for efficient photosynthesis in Arabidopsis. The Plant Cell **15**, 1817-1832.

Pierre M, Traverso JA, Boisson B, Domenichini S, Bouchez D, Giglione C, Meinnel T. 2007. N-Myristoylation regulates the SnRK1 pathway in Arabidopsis. The Plant Cell **19**, 2804-2821.

Pletnev PI, Shulenina O, Evfratov S, Treshin V, Subach MF, Serebryakova MV, Osterman IA, Paleskava A, Bogdanov AA, Dontsova OA, Konevega AL, Sergiev PV. 2022. Ribosomal protein S18 acetyltransferase RimI is responsible for the acetylation of elongation factor Tu. The Journal of Biological Chemistry **298**, 101914.

Poulard C, Treilleux I, Lavergne E, Bouchekioua-Bouzaghou K, Goddard-Léon S, Chabaud S, Trédan O, Corbo L, Le Romancer M. 2012. Activation of rapid oestrogen signalling in aggressive human breast cancers. EMBO Molecular Medicine **4**, 1200-1213.

Prus G, Hoegl A, Weinert BT, Choudhary C. 2019. Analysis and interpretation of protein post-translational modification site stoichiometry. Trends in Biochemical Sciences **44**, 943-960.

Qi Q, Rajala RV, Anderson W, Jiang C, Rozwadowski K, Selvaraj G, Sharma R, Datla R. 2000. Molecular cloning, genomic organization, and biochemical characterization of myristoyl-CoA:protein N-myristoyltransferase from *Arabidopsis thaliana*. The Journal of Biological Chemistry **275**, 9673-9683.

Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC. 2012. Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana. Nature Communications **3**, 686.

Raven JA. 2013. Rubisco: still the most abundant protein of Earth? The New Phytologist **198**, 1-3.

Redman K, Rubenstein PA. 1981. NH2-terminal processing of Dictyostelium discoideum actin in vitro. The Journal of Biological Chemistry **256**, 13226-13229.

Redman KL, Martin DJ, Korn ED, Rubenstein PA. 1985. Lack of NH2-terminal processing of actin from Acanthamoeba castellanii. The Journal of Biological Chemistry **260**, 14857-14861.

Reid DW, Nicchitta CV. 2015. Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nature Reviews Molecular Cell Biology

16, 221-231.

Renna L, Stefano G, Majeran W, Micalella C, Meinnel T, Giglione C, Brandizzi F. 2013. Golgi traffic and integrity depend on N-myristoyl transferase-1 in Arabidopsis. The Plant Cell **25**, 1756-1773.

Richter S, Lamppa GK. 1998. A chloroplast processing enzyme functions as the general stromal processing peptidase. Proceedings of the National Academy of Sciences of the United States of America **95**, 7463-7468.

Richter S, Lamppa GK. 2002. Determinants for removal and degradation of transit peptides of chloroplast precursor proteins. The Journal of Biological Chemistry **277**, 43888-43894.

Richter S, Zhong R, Lamppa G. 2005. Function of the stromal processing peptidase in the chloroplast import pathway. Physiologia Plantarum **123**, 362-368.

Ross S, Giglione C, Pierre M, Espagne C, Meinnel T. 2005. Functional and developmental impact of cytosolic protein N-terminal methionine excision in *Arabidopsis*. Plant Physiology **137**, 623-637.

Rouhier N, Vieira Dos Santos C, Tarrago L, Rey P. 2006. Plant methionine sulfoxide reductase A and B multigenic families. Photosynthesis Research **89**, 247-262.

Rowland E, Kim J, Bhuiyan NH, van Wijk KJ. 2015. The Arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. Plant Physiology **169**, 1881-1896.

Rubenstein PA, Martin DJ. 1983. NH2-terminal processing of Drosophila melanogaster actin. Sequential removal of two amino acids. The Journal of Biological Chemistry **258**, 11354-11360.

Schilling S, Stenzel I, von Bohlen A, Wermann M, Schulz K, Demuth HU, Wasternack C. 2007. Isolation and characterization of the glutaminyl cyclases from Solanum tuberosum and Arabidopsis thaliana: implications for physiological functions. Biological Chemistry **388**, 145-153.

Schilling S, Wasternack C, Demuth HU. 2008. Glutaminyl cyclases from animals and plants: a case of functionally convergent protein evolution. Biological Chemistry **389**, 983-991.

Serero A, Giglione C, Meinnel T. 2001. Distinctive features of the two classes of eukaryotic peptide deformylases. Journal of Molecular Biology **314**, 695-708.

Sheff DR, Rubenstein PA. 1992. Isolation and characterization of the rat liver actin N-acetylaminopeptidase. The Journal of Biological Chemistry **267**, 20217-20224.

Sheiner L, Soldati-Favre D. 2008. Protein trafficking inside Toxoplasma gondii. Traffic **9**, 636-646.

Shipman-Roston RL, Ruppel NJ, Damoc C, Phinney BS, Inoue K. 2010. The significance of protein maturation by plastidic type I signal peptidase 1 for thylakoid development in Arabidopsis chloroplasts. Plant Physiology **152**, 1297-1308.

Stahl A, Moberg P, Ytterberg J, Panfilov O, Brockenhuus Von Lowenhielm H, Nilsson F, Glaser E. 2002. Isolation and identification of a novel mitochondrial metalloprotease (PreP) that degrades targeting presequences in plants. The Journal of Biological Chemistry **277**, 41931-41939.

Suazo KF, Park KY, Distefano MD. 2021. A not-so-ancient grease history: click chemistry and protein lipid modifications. Chemical Reviews **121**, 7178-7248.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. Journal of Experimental Botany **59**, 1515-1524.

Tamura T, Asakura T, Uemura T, Ueda T, Terauchi K, Misaka T, Abe K. 2008. Signal peptide peptidase and its homologs in Arabidopsis thaliana– plant tissue-specific expression and distinct subcellular localization. The FEBS Journal **275**, 34-43.

Tarrago L, Laugier E, Rey P. 2009. Protein-repairing methionine sulfoxide reductases in photosynthetic organisms: gene organization, reduction mechanisms, and physiological roles. Molecular Plant **2**, 202-217.

Teixeira PF, Kmiec B, Branca RM, Murcha MW, Byzia A, Ivanova A, Whelan J, Drag M, Lehtiö J, Glaser E. 2017. A multi-step peptidolytic cascade for amino acid recovery in chloroplasts. Nature Chemical Biology **13**, 15-17.

Tonkin CJ, Kalanon M, McFadden GI. 2008. Protein targeting to the malaria parasite plastid. Traffic **9**, 166-175.

Traverso JA, Giglione C, Meinnel T. 2013a. High-throughput profiling of N-myristoylation substrate specificity across species including pathogens. Proteomics **13**, 25-36.

Traverso JA, Micalella C, Martinez A, Brown SC, Satiat-Jeunemaitre B, Meinnel T, Giglione C. 2013b. Roles of N-Terminal fatty acid acylations in membrane compartment partitioning: Arabidopsis h-type thioredoxins as a case study. The Plant Cell **25**, 1056-1077.

Tripathi LP, Sowdhamini R. 2006. Cross genome comparisons of serine proteases in Arabidopsis and rice. BMC Genomics **7**, 200.

Tsiatsiani L, Stael S, Van Damme P, Van Breusegem F, Gevaert K. 2014. Preparation of Arabidopsis thaliana seedling proteomes for identifying metacaspase substrates by N-terminal COFRADIC. Methods in Molecular Biology **1133**, 255-261.

Tsiatsiani L, Timmerman E, De Bock PJ, Vercammen D, Stael S, van de Cotte B, Staes A, Goethals M, Beunens T, Van Damme P, Gevaert K, Van Breusegem F. 2013. The Arabidopsis metacaspase9 degradome. The Plant Cell **25**, 2831-2847.

Tsuji A, Fujisawa Y, Mino T, Yuasa K. 2011. Identification of a plant aminopeptidase with preference for aromatic amino acid residues as a novel member of the prolyl oligopeptidase family of serine proteases. Journal of Biochemistry **150**, 525-534.

Van Damme P, Evjenth R, Foyn H, Demeyer K, De Bock PJ, Lillehaug JR, Vandekerckhove J, Arnesen T, Gevaert K. 2011. Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. Molecular & Cellular Proteomics **10**, M110 004580.

Van Damme P, Lasa M, Polevoda B, Gazquez C, Elosegui-Artola A, Kim DS, De Juan-Pardo E, Demeyer K, Hole K, Larrea E, Timmerman E, Prieto J, Arnesen T, Sherman F, Gevaert K, Aldabe R. 2012. N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proceedings of the National Academy of Sciences of the United States of America **109**, 12449-12454.

van Dooren GG, Striepen B. 2013. The algal past and parasite present of the apicoplast. Annual Review of Microbiology **67**, 271-289.

van Dooren GG, Su V, D'Ombrain MC, McFadden GI. 2002. Processing of an apicoplast leader sequence in Plasmodium falciparum and the identification of a putative leader cleavage enzyme. The Journal of Biological Chemistry **277**, 23612-23619.

Varland S, Vandekerckhove J, Drazic A. 2019. Actin post-translational modifications: the Cinderella of cytoskeletal control. Trends in Biochemical Sciences **44**, 502-516.

Varshavsky A. 2019. N-degron and C-degron pathways of protein degradation. Proceedings of the National Academy of Sciences **116**, 358-366.

Venne AS, Solari FA, Faden F, Paretti T, Dissmeyer N, Zahedi RP. 2015. An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in Arabidopsis thaliana. Proteomics **15**, 2458-2469.

Vicente J, Mendiondo GM, Pauwels J, Pastor V, Izquierdo Y, Naumann C, Movahedi M, Rooney D, Gibbs DJ, Smart K, Bachmair A, Gray JE, Dissmeyer N, Castresana C, Ray RV, Gevaert K, Holdsworth MJ. 2019. Distinct branches of the N-end rule pathway modulate the plant immune response. The New Phytologist **221**, 988-1000.

Wakaguri H, Suzuki Y, Sasaki M, Sugano S, Watanabe J. 2009. Inconsistencies of genome annotations in apicomplexan parasites revealed by 5'-end-one-pass and full-length sequences of oligo-capped cDNAs. BMC Genomics **10**, 312.

Walling LL. 2006. Recycling or regulation? The role of amino-terminal modifying enzymes. Current Opinion in Plant Biology **9**, 227-233.

Weits DA, Giuntoli B, Kosmacz M, Parlanti S, Hubberten HM, Riegler H, Hoefgen R, Perata P, van Dongen JT, Licausi F. 2014. Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nature Communications **5**, 3425.

Weits DA, Kunkowska AB, Kamps NCW, Portz KMS, Packbier NK, Nemec Venza Z, Gaillochet C, Lohmann JU, Pedersen O, van Dongen JT, Licausi F. 2019. An apical hypoxic niche sets the pace of shoot meristem activity. Nature **569**, 714-717.

Weng SSH, Demir F, Ergin EK, Dirnberger S, Uzozie A, Tuscher D, Nierves L, Tsui J, Huesgen PF, Lange PF. 2019. Sensitive determination of proteolytic proteoforms in limited microscale proteome samples. Molecular & Cellular Proteomics **18**, 2335-2347.

Wetzel CM, Harmacek LD, Yuan LH, Wopereis JL, Chubb R, Turini P. 2009. Loss of chloroplast protease SPPA function alters high light acclimation processes in Arabidopsis thaliana L. (Heynh.). Journal of Experimental Botany **60**, 1715-1727.

White MD, Klecker M, Hopkinson RJ, Weits DA, Mueller C, Naumann C, O'Neill R, Wickens J, Yang J, Brooks-Bartlett JC, Garman EF, Grossmann TN, Dissmeyer N, Flashman E. 2017. Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. Nature Communications **8**, 14690.

Wilson RJ, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH. 1996. Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. Journal of Molecular Biology **261**, 155-172.

Xu F, Huang Y, Li L, Gannon P, Linster E, Huber M, Kapos P, Bienvenut W, Polevoda B, Meinnel T, Hell R, Giglione C, Zhang Y, Wirtz M, Chen S, Li X. 2015. Two N-terminal acetyltransferases antagonistically regulate the stability of a nod-like receptor in Arabidopsis. The Plant Cell **27**, 1547-1562.

Yamaguchi K, Subramanian AR. 2000. The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). The Journal of Biological Chemistry **275**, 28466-28482.

Yamauchi Y, Ejiri Y, Toyoda Y, Tanaka K. 2003. Identification and biochemical characterization of plant acylamino acid-releasing enzyme. Journal of Biochemistry **134**, 251-257.

Ying Z, Mulligan RM, Janney N, Houtz RL. 1999. Rubisco small and large subunit N-methyltransferases. Bi- and mono-functional methyltransferases that methylate the small and large subunits of Rubisco. The Journal of Biological Chemistry **274**, 36750-36756.

Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Molecular Biology and Evolution **22**, 1299-1308.

Yoshida S, Ito M, Callis J, Nishida I, Watanabe A. 2002. A delayed leaf senescence mutant is defective in arginyl-tRNA:protein arginyltransferase, a component of the N-end rule pathway in *Arabidopsis*. The Plant Journal **32**, 129-137.

Zahn-Zabal M, Michel PA, Gateau A, Nikitin F, Schaeffer M, Audot E, Gaudet P, Duek PD, Teixeira D, Rech de Laval V, Samarasinghe K, Bairoch A, Lane L. 2020. The neXtProt knowledgebase in 2020: data, tools and usability improvements. Nucleic Acids Research

48, D328-d334.

Zhang H, Deery MJ, Gannon L, Powers SJ, Lilley KS, Theodoulou FL. 2015. Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots. Proteomics **15**, 2447-2457.

Zhang H, Gannon L, Hassall KL, Deery MJ, Gibbs DJ, Holdsworth MJ, van der Hoorn RAL, Lilley KS, Theodoulou FL. 2018. N-terminomics reveals control of Arabidopsis seed storage proteins and proteases by the Arg/N-end rule pathway. The New Phytologist **218**, 1106-1126.

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. 2008. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PloS One **3**, e1994.

587

588 **TABLES**

589 **Table 1. The amino terminal proteolytic processing machinery of plastid-containing organisms**

Aminopeptidase	BRENDA category	Entry TAIR	Gene Name UniProt	Cleavage specificity	Locali-zation	References	Strame-nopiles	Rhiza-ria	Apicom-plexa
Organellar									
Presequence processing									
Chloroplast processing peptidase	3.4.21.89	At3g24590	PLSP1	Processing of some plastid-targeted proteins including OE33 & Toc75	C	(Inoue *et al.*, 2005; Shipman-Roston *et al.*, 2010)	-		
Presequence protease 1	3.4.24.-	At3g19170	PREP1	X(10-65)RR (LFY)/(SAV)(STA)	C/M	(Huang *et al.*, 2009; Kmiec and Glaser, 2012; Stahl *et al.*, 2002)	-		
Presequence protease 2		At1g49630	PREP2						
Stromal processing peptidase	3.4.21.102	At5g42390	SPP	Presequence rich in S and T	C	(Richter and Lamppa, 1998, 2002; Richter *et al.*, 2005)	Naga_100079g9	-	TGME49_035680
Serine protease SPPA, chloroplastic	3.4.21.-	At1g73990	SPPA	Degradation of antenna and PSII	C	(Lensch *et al.*, 2001; Wetzel *et al.*, 2009)	Naga_100001g124	B1X5C9	
Aminopeptidase P, prolidase	3.4.11.26	At4g29490		?RX(FY)/(AS)??	C	(Aryal *et al.*, 2014; Huang *et al.*, 2015)			
Mitochondrial-processing peptidase subunit alpha	3.4.24.64	At3g16480	MMPA 2	Might have no protease activity (lacks one conserved zinc-binding site)	C/M	(Baudisch and Klösgen, 2012)	Naga_100011g96, Naga_100003g161		
		At1g51980	MMPA 1		M	(Carrie *et al.*, 2015)	Naga_100066g1		TGME49_002680
Mitochondrial-processing peptidase subunit beta		At3g02090	MPPB		M		NGA_016780, Naga_100054g10		

Intermediate cleaving peptidase 55, Xaa-Pro dipeptidase	3.4.13.9	At1g09300	ICP55	RX(FY)/(AS)	M	(Carrie *et al.*, 2015; Huang *et al.*, 2015)		B1X3Z2	
Mitochondrial intermediate peptidase	3.4.24.-	At5g51540	OCT1	S(AS)/A	M	(Carrie *et al.*, 2015)	Naga_100005g90		
Peptide deformylation									
Peptide deformylase 1B	3.5.1.88	At5g14660	PDF1B	f/MX	C/M	(Dirk *et al.*, 2001; Giglione *et al.*, 2000b)	Naga_100009g109	B1X553	TGME49_016700
Peptide deformylase 1A		At1g15390	PDF1A	f/MX	C/M				
N-terminal Methionine cleavage									
Methionine aminopeptidase 1C	3.4.11.18	At1g13270	MAP1B *	M/(AGPTVR))(^P)	C	(Giglione *et al.*, 2000b)	Naga_100084g2	B1X3G2	TGME49_057730, TGME49_211330
Methionine aminopeptidase 1B		At3g25740	MAP1C *	M/(AGPTV))	C/M				
Methionine aminopeptidase 1D		At4g37040	MAP1D	M/(AGPTVN)	C/M				
N-terminal processing mono- or di- aminopeptidases									
Puromycin-sensitive aminopeptidase	3.4.11.14	At1g63770	PSA	R>AGILMK/X	C/M	(Teixeira *et al.*, 2017)	Naga_100111g8	g3466.t1	
Leucine aminopeptidase 3	3.4.11.5	At4g30910	LAP3	ML>FWY/X	C/M	(Teixeira *et al.*, 2017)	Naga_100065g23		
Leucine aminopeptidase 2		At4g30920	LAP2		C/M				
Aspartyl aminopeptidase	3.4.11.7	At5g04710	DAP2	(DE)/X	C/M	(Park *et al.*, 2017; Teixeira *et al.*, 2017)	Naga_100236g6		
Aminopeptidase P2	3.4.11.9	At3g05350	AMPP2	X/P?	C? cytosol(Aryal *et al.*, 2014)	(Ferro *et al.*, 2010; Murphy *et al.*, 2002)	Naga_100013g76		
Tyrosyl aminopeptidase	3.4.11.x	At5g36210	MAB16	(YFM)/X	C	(Tsuji *et al.*, 2011)	NGA_0474200		TGME49_246800

Prolyl endopeptidase	3.4.21.-	At1g76140		X(P>A)/X	C/M	(Tripathi and Sowdhamini, 2006)	Naga_100093g9		
Acylamino-acid-releasing enzyme	3.4.19.1	At4g14570.1	AARE	AcX/X	C	(Yamauchi *et al.*, 2003; Zybailov *et al.*, 2008)	-		
Glutamyl endopeptidase	3.4.21.-	At2g47390	CGEP	PE/X	C	(Bhuiyan *et al.*, 2020; Forsberg *et al.*, 2005)	-		
Cytosolic									
Presequence processing									
Signal peptide peptidase	3.4.23.B2	At2g03120	SPP	Cleaves type II membrane signal peptides in the hydrophobic plane of the membrane	ER	(Tamura *et al.*, 2008)	Naga_100019g45		TGME49_037150
Signal peptide peptidase-like 1		At4g33410	SPPL1		Endosome		Naga_100077g7		
Signal peptide peptidase-like 2		At1g63690	SPPL2						
Signal peptide peptidase-like 3		At2g43070	SPPL3						
Signal peptide peptidase-like 4		At1g01650	SPPL4						
Signal peptide peptidase-like 5		At1g05820	SPPL5						
Signal peptidase I complex subunit SEC11	3.4.21.89	At1g52600	SPC18	Removes signal peptides from nascent proteins as they are translocated into the lumen of the endoplasmic reticulum	ER MB		Naga_100086g8		
		At3g15710							
Signal peptidase complex subunit 3A		At3g05230	SPC22				Naga_100105g6		
Signal peptidase complex subunit 3B		At5g27430	SPC23						
Signal peptidase complex subunit 2		At2g39960	SPC25				Naga_100222g1		
		At4g04200							

Signal peptidase complex subunit 1		At2g22425	SPC12				-		
N-terminal Methionine cleavage									
Methionine aminopeptidase 1A	3.4.11.18	At2g45240	MAP1A	M/(AGPTVR))(^P)		(Giglione *et al.*, 2000b; Ross *et al.*, 2005)	Naga_100013g54		TGME49_048850
Methionine aminopeptidase 2A		At2g44180	MAP2A	M/(AGPTV))			Naga_100004g92		TGME49_105460
Methionine aminopeptidase 2B		At3g59990	MAP2B						
N-terminal processing mono- or di- aminopeptidases									
Pyroglutamyl peptidase	3.4.19.-	At1g56700	PGP1	Glp[£]/peptide			No but in most Ochrophyta		TGME49_268280
		At1g23440	PGP2						
Aspartyl aminopeptidase	3.4.11.7	At5g60160	DAP1					g40631.t1	
X-Pro dipeptidase	3.4.13.9	At4g29500			C		Naga_100025g65		
Aminopeptidase M1	3.4.11.2	At4g33090	APM1	A/X	C		Naga_100070g24	g54899.t1	
Aminopeptidase P1	3.4.11.9	At4g36760	AMPP1	Y/X	Mb	(Murphy *et al.*, 2002)	Naga_100013g76		
Leucine aminopeptidase 1	3.4.11.-	At2g24200	AMPL1				Naga_100005g129		
Leucine aminopeptidase	3.4.11.-	At5g13520	LKHA4	cleaves only di or tri peptides			Naga_100009g11		
Dipeptidyl peptidase IV	3.4.21.-	At5g24260	MOP9.8	XP/Y	C	(Tripathi and Sowdhamini, 2006)	-	g69768.t1	
Prolyl endopeptidase	3.4.21.-	At1g20380			C		Naga_100093g9		
Acylamino-acid-releasing enzyme	3.4.19.1	At4g14570	AARE	AcX/X	C	(Aryal *et al.*, 2017; Nakai *et al.*, 2012)	Naga_100324g4		

Sequence information is from UniProt for *A. thaliana*. Stremenopiles is exemplified with sequences from *Nannochloropsis gaditana* and the whole phylum was searched if matching was unsuccessful; the Apicomplexa phylum was searched through *Toxoplasma gondii* ME49.

[&] Further information and early bibliographic references can be found in (Walling, 2006)

[£] Glp is pyroglutamate
[*] inverted names in TAIR

595 **Table 2. Oxidative and addition processing machinery of plastid-containing organisms.**

Protein N-terminal Modification	Protein function	Name	BRENDA category	Entry TAIR	Gene Name TAIR	Targets	Localization	References	*Nannochloropsis gaditana*[£]	*Paulinella* spp	*Toxoplasma gondii*
Formylation	Methionyl-tRNA formyltransferase		2.1.2.9	At1g66520	pde194	Met-$tRNA_f^{Met}$	Plastid/ Mitochondria	(Bryant *et al.*, 2011)	jgi|Nangad1|4821|rna8625	A0A1S6YJ55	TGME49_235640
Arginylation	Arginyl-tRNA-protein transferase	ATE1	2.3.2.8	At5g05700	DLS1	ERF-VII transcription factors	Cytosol	(Graciet *et al.*, 2009; White *et al.*, 2017)	Naga_100523g2	-	-
		ATE2		At3g11240					Naga_100001g75		
Monomethylation	Protein methyltransferase	NTM1	2.1.1.244	At5g44450		[APS]-P-K	Cytosol		Naga_100644g2	-	
		?		?		RPL2(AIV) , RPL16 (MLS)	Plastid	(Yamaguchi and Subramanian, 2000)			TGME49_201720
		RBCMT		At1g14030	RBCMT	RbcS (MQV)	Plastid	(Ying *et al.*, 1999) not checked in AT	Naga_100013g90	-	-
N-myristoylation	N-myristoyltransferase	NMT1	2.3.1.97	At5g57020		GXXXSX	Cytosol	(Boisson *et al.*, 2003; Pierre *et al.*, 2007; Qi *et*	Naga_100004g141	No but homologs do occur in Cercozoa, (Giglione and Meinnel,	TGME49_209160
		NMT2		At2g44170			Cytosol				

								al., 2000)		2022)	
Acetylation	NatA complex	NAA10	2.3.1.255	At5g13780	ARD1	[ACGSTV]	Cytosol	(Linster *et al.*, 2015; Miklánková *et al.*, 2022)	Naga_100001g191	XP_028065450.1	TGME49_219760
		HYPK		At3g06610					Naga_100133g15	-	-
		NAA15		At1g80410					Naga_100115g20	XP_024387243.1	TGME49_223710
	NatB complex	NAA20	2.3.1.254	At1g03150	NAT5	M[DENQ]		(Huber *et al.*, 2020)	Naga_100011g91	XP_008876485.1	TGME49_289900
		NAA25		At5g58450					jgi|Nanoce1779_2|633493*	GBB94039.1	TGME49_203750
	NatC complex	NAA30	2.3.1.256	At2g38130	MAK3	M[KL]		(Pesaresi *et al.*, 2003)	Naga_100015g22	CEP01279.1	TGME49_259070
		NAA35		At2g11000	MAK10				Naga_100166g2	XP_004338532.1	-
	NatD	NAA40	2.3.1.257	At1g18335		S(H2A)			Naga_100019g18, Naga_100074g15	Yes but not retrieved	TGME49_230060
	NatE	NAA50	2.3.1.-	At5g11340				(Armbruster *et al.*, 2020)	Naga_100126g8	-	TGME49_260010
	NatH	NAA60		At5g16800				(Linster *et al.*, 2020)	jgi|Nangad1|5987|rna9419	-	TGME49_260550
	NatG	NAA70		At4g28030			Plastid	(Bienvenut *et al.*, 2020; Dinh *et al.*, 2015)	Naga_100009g37	-	-
		NAA90		At1g26220, At1g32070, At4g19985		NSI			-	APP88668.1, YP_009530844.1	-
Deamidation	Nt-Asn deamidase	NTAN1	3.5.1.121	At2g44420			Cytosol	(Vicente *et al.*, 2019)	Naga_100337g5	-	-

	Nt-Gln deamidase	NTAQ1	3.5.1.12 2	At2g41760							-
Cysteine oxidation	Nt-Cys oxidase	PCO4	1.13.11. 20	At2g42670			Cytosol	(Weits *et al.*, 2014)	-	-	-
		PCO5		At3g58670							
		PCO1		At5g15120							
		PCO2		At5g39890							
Methionine reduction	Peptide-methionine (S)-S-oxide reductase	MSRA1	1.8.4.11	At5g61640	MRSA		Cytosol	(Rouhier *et al.*, 2006)	Naga_100126g3	-	TGME49_225470
		MSRA2		At5g07460			Cytosol				
		MSRA3		At5g07470			Cytosol		Naga_100404g4		
		MSRA4		At4g25130			Plastid		Naga_101420g2		
		MSRA5		At2g18030			Secretory				
	Peptide-methionine (R)-S-oxide reductase	MSRB1	1.8.4.12	At1g53670	MRSB		Plastid		Naga_100094g2	-	TGME49_114000
		MSRB2		At4g21860			Plastid				
		MSRB3		At4g04800			Secretory				
		MSRB4		At4g04810			Cytosol				
		MSRB5		At4g04830			Cytosol				
		MSRB6		At4g04840			Cytosol				
		MSRB7		At4g21830			Cytosol				
		MSRB8		At4g21840			Cytosol				
		MSRB9		At4g21850			Cytosol				

* only retrieved in *Nannochloropsis oceanica*, not in *N. gaditana*

£ searches performed first at NCBI and, if unsuccessful, further processed at JGI (https://phycocosm.jgi.doe.gov/pages/blast-query.jsf?db=eustigmatophyta); the NCBI ID is indicated if retrieved and otherwise the JGI entry code.

599 **Table 3. Phylogeny of RbcL type 1 is exemplified by differences at the N-termini and the associated NPTMs.**

Domain	Phylum	Order	Plastid acquisition	Species	Type	Lineage	Unprocessed translated sequence	N-terminal sequence	References
Eukaryota	Rhizaria	Cercozoa	Primary	Paulinella micropora	Green	1A	MS-KKYDAGVKEYRDT-YWTPDYIPLDTDLLACFKCTGQE	S2 or AcS2	
				Paulinella chromatophora			MS-KKYDAGVKEYRDT-YWTPDYIPLDTDLLACFKCTGQE		
Bacteria	Cyanobacteria	Cyanophyceae	-	Prochlorococcus marinus			MS-KKYDAGVKEYRDT-YWTPEYVPLDTDLLACFKCTGQE		
				Synechococcus sp.BL107			MS-KKYDAGVKEYRDT-YWTPDYVPLDTDLLACFKCTGQE		
	Alphaproteo-bacteria	Rhodobacterales		Rhodovulum sulfidophilum			MA-KTYDAGVKDYRST-YWEPDYTLKDSDILAVFKVIPQP	A2	
				Rhodobacter capsulatus			MAAKTYDAGVKDYRSI-YWEPQYQVKDSDILAVFKVVPQP		
Bacteria	Cyanobacteria	Cyanophyceae	-	Synechocystis sp. PCC6803		1B	MVQAKAGFKAGVQDYRLT-YYTPDYTPKDTDLLACFRMTPQP	V2	
				Synechococcus elongatus PCC118			MPKTQSAAGYKAGVKDYKLT-YYTPDYTPKDTDLLAAFRFSPQP	P2	
				Synechococcus sp.PCC7335			MSYSQTQSKSGAGYDAGVQDYRLT-YYAPDYTPRDTDILAAFRMTPQP	S2	
				Nostoc sp.PCC7120			MSYAQTKTQTKSGYKAGVQDYRLT-YYTPDYTPKDTDILAAFRVTPQP	S2	(Flecken *et al.*, 2020)
Eukaryota	Plantae/ Archaeplastida	Glaucophyta	Primary	Cyanophora paradoxa			MS-SQARTQTRAGFKAGVKDYRLT-YYTPEYTPKETDILAAFRMTPQP	S2	
				Glaucocystis incrassate			MS-SQARTQTVAGFKAGVKDYRLN-YYTPDYKPKETDILAAFRMTPQP	S2	
				Gloeochaete wittrockiana			MP-SQSRTQTASGFKAGVKDYRLN-YYTPEYTPKETDILAAFRMTPQP	P2	
		Chlorophyta		Ostreococcus tauri			MA-PQTETKTGTGFQAGVKDYRLT-YYTPDYQVKETDILAAFRMTPQP	AcP3	
				Chlamydomonas reinhardtii			MV-PQTETKAGAGFKAGVKDYRLT-YYTPDYVVRDTDILAAFRMTPQL		(Houtz *et al.*, 1992)
				Physcomitrium patens			MS-PRPEIKAGVGFKAGVKDYRLT-YYTPDYQTKDTDILAAFRMTPQP		
				Marchantia polymorpha			MS-PQTETKAGVGFKAGVKDYRLT-YYTPDYETKDTDILAAFRMTPQP		(Houtz *et al.*, 1992)
				Triticum aestivum			MS-PQTETKASVGFKAGVKDYKLT-YYTPEYETKDTDILAAFRVTPQP		(Houtz *et al.*, 1992)

				Nicotiana tabacum			MS-PQTETKASVGFKAGVKEYKLT-YYTPEYQTKDTDILAAFRVTPQP		(Houtz *et al.*, 1992)
				Arabidopsis thaliana			MS-PQTETKASVGFKAGVKEYKLT-YYTPEYETKDTDILAAFRVTPQP		(Bienvenut *et al.*, 2012)
	Excavata	Euglenales	Secondary	Euglena gracilis			MS-PQTETKTGAGFKAGVKDYRLT-YYTPDYQVSETDILAAFRMTPQP		
	Rhizaria	Cercozoa	Secondary	Bigelowiella natans			M-IFAKKTE-AAGFKAGVKDYRLT-YYTPDYKVKDTDILAAFRVTPQP	M1	
Bacteria	Alphaproteo-bacteria	Hyphomicrobiales	-	Rhodopseudomonas palustris	Red	1C	MNESVTI---RGKDRYKSGVMEYKKMGYWEPDYEPKDTDIIALFRVTPQD	M1	
		Rhodobacterales		Rhodobacter sphaeroides			MDT--KTTEIKGKERYKAGVLKYAQMGYWDGDYVPKDTDVLALFRITPQE	M1	
				Polymorphum gilvum			MDA--KMKEIKGKERYKAGVLKYAQMGYWDGDYEPKDTDVIALFRITPQD	M1	
	Betaproteobacteria	Burkholderiales		Ralstonia eutropha			MNAPETIQAKPRK-RYDAGVMKYKEMGYWDGDYVPKDTDVLALFRITPQD	M1	
Eukaryota	SAR	Stramenopiles	Secondary	Thalassiosira pseudonana		1D	MSQSVSERTRIKSDRYESGVIPYAKMGYWDAAYTVKDTDVLALFRITPQP	AcS2	(Huesgen *et al.*, 2013)
				Odontella sinensis			MSQSVSERTRIKSDRYESGVIPYAKMGYWDASYTVQDTDVLALFRITPQP		
				Diatoma tenuis			MSQSVTERTRIKSDRYESGVIPYAKMGYWDASYAVKTTDVLALFRITPQP		
				Phaeodactylum tricornutum			MSQSVSERTRIKSDRYESGVIPYAKMGYWDAAYAVKNTDVLALFRITPQP		
				Olisthodiscus luteus			MSNNVYERNRIKNDRYESGVIPYAKMGYWDANYSVKDTDLLALFRLTPQP		
	CCTH	Cryptophyta		Guillardia theta			MSQSVESRTRIKNERYESGVIPYAKMGYWDADYVIKDTDVLAMFRMTPQK		
				Cryptomonas paramecium			MSQSVETRTRIKNERYESGVIPYAKMGYWDADYSIKDTDILAMFRMTPQK		
		Haptophyta		Emiliania huxleyi			MSQAVESRTRIKSERYESGVIPYAKMGYWDPEYVIKDTDILALFRCTPQP		
	Plantae/ Archaeplastida	Rhodophyta	Primary	Pyropia yezoensis			MSQSVESRTRIKSERYESGVIPYAKMGYWDADYVIKETDILALFRITPQP		
				Rhodochaete parvula			MSQSVEERTRIKNERYESGVIPYARMGYWDADYVIKDTDVLALFRISPQP		
				Cyanidioschyzon merolae			MAQSVQERTRLKNKRYESGVIPYAKMGYWNPNYVVKDSDMLALFRVTPQP	AcA2	
				Cyanidium caldarium			MAQSVQERTRLKNKRYESGVIPYAKMGYWDPNYVVKDTDILALFRVTPQP	A2	

Table 4. Cytosolic actin isoforms from different organisms, their N-terminus, and the N-terminal machinery involved.

Clade	Infra-kingdom	Species	Protein name	Alternative name	UniProtKB entry	N-terminal sequence	Matured N-terminus	Matur-ation class	MetAP cleavage	Nat(s) involved		ANAP	Reference
Unikonta	Metazoa	*Homo sapiens*	ACTC	α-cardiac	P68032	MCDDEE	AcC2	IIa	Y	A	-	N	(Van Damme *et al.*, 2012; Varland *et al.*, 2019)
			ACTS	α-skeletal	P68133	MCDEDE	AcC2		Y	A	-	N	
			ACTA	α-smooth	P62736	MCEEED	AcE3		Y	A	H	Y	
			ACTB	β-cyto	P60709	MDDD	AcM1, AcD2 & AcD3	Ib	N	B	H	Y	
			ACTG	γ-cyto	P63261	MEEE	AcM1, AcE2 & AcE3		N	B	H	Y	
			ACTH	γ-smooth	P63267	MCEEE	AcE3	IIb	Y	A	H	Y	
		Drosophila melanogaster	ACT1		P10987	MCDEE	AcD3		Y	A	H	Y	(Rubenstein and Martin, 1983)
	Fungi	*Schizo-Saccharomyces pombe*	ACT1		P10989	MEEE	AcM1	Ia	N	B	-	N	(Cook *et al.*, 1991)
		Candidat albicans	ACT		A0A1D8PFR4	MDGEE	AcM1		N	B	-	N	(Cook *et al.*, 1991)
		Saccharomyces cerevisiae	ACT1		P60010	MDSE	AcM1		N	B	-	N	(Van Damme *et al.*, 2012)
	Amoebozoa	*Acanthamoeba castellanii*	ACT1		P02578	MGDE	AcG2	IIa	Y	A	-	N	(Hernández-Cuevas *et al.*, 2019; Redman *et al.*, 1985)
		Dictyostelium discoideum	ACT1		P07830	MDGED	AcD2	Ib	N	B	H	Y	(Redman and Rubenstein, 1981)
Bikonta	Archaeplastida	*Arabidopsis thaliana*	ACT1	At2g37620	P0CJ46	MADGED	AcA2	IIa	Y	A	-	N	NterDB
			ACT2	At3g18780	Q96292	MAEADD	AcA2		Y	A	-	N	

			ACT3	At3g53750	P0CJ47	MADGED	AcA2		Y	A	-	N	
			ACT4	At5g59370	P53494	MADGED	AcA2		Y	A	-	N	
			ACT7	At5g09810	P53492	MADGED	AcA2		Y	A	-	N	
			ACT8	At1g49240	Q96293	MADADD	AcA2		Y	A	-	N	
			ACT11	At3g12110	P53496	MADGED	AcA2		Y	A	-	N	
			ACT12	At3g46520	P53497	MADGED	AcA2		Y	A	-	N	
	Rhizaria	*Paulinella chromate-phera*	ACT1			MAEEEEEE	AcE3	IIb	Y	A	H?	Y	(Oberleitner *et al.*, 2022)
	Alveolata	*Plasmodium berghei*	ACT1		A0A509AU65	MGDEE	AcG2	IIa	Y	A	-	N	(Nyonda *et al.*, 2022)
		Toxoplasma gondii	ACT1		P53476	MADEE	AcA2 & AcD3	IIb	Y	A	H?	Y	
	Stramenopiles	*Thalassiosira pseudonana*	ACT1		B8CFV4	MADEED	AcA2	IIa	Y	A	-	N	(Huesgen *et al.*, 2013)

NterDB is available at https://n-terdb.i2bc.paris-saclay.fr/.

BOX

Box 1: The plastid-encoded N-terminal processing machinery and its progressive transfer to the nuclear genome throughout evolution

Here we illustrate with the example of the N-terminal processing machinery of plastid-containing organisms how gene transfer might have proceeded progressively from plastid to nucleus during evolution. We use the two subunits of RuBisCO (RbcL and RbcS) as markers of the migration of major photosynthetic functions borne by several subunits now encoded in both genomes (this also includes photosystems). In the top left of **Figure Box 1**, we list all enzymes of the early modification machinery also occurring in cyanobacteria - namely FMT, PDF, MetAP, PepP, RimIJ, and NAA90 - and their localization during three major stages of evolution. RimIJL corresponds to the only NAT occurring in bacteria and modifying only a few proteins including EFTu, SecB, and ribosomal proteins L7 and S5 (Bienvenut *et al.*, 2015; Pletnev *et al.*, 2022). The corresponding timescale is shown in the bottom green panel. This illustrates how the plastid content decreased from a couple of thousand genes to 850 in Paulinellidae, 150-200 in red algae, and 90 or so in land plants. In the right blue panel, the type of evolution is indicated for each gene. We indicate in this figure how major genes involved in early N-terminal modifications were lost very late in evolution. The most remarkable is certainly NAA90, which was transferred to the nucleus contemporary to RbcS. This analysis also suggests that NAA70 might have evolved later directly from the nuclear version of NAA70. This might explain why only NAA70 has been retrieved in Euglenozoa and diatoms even though NAA70 was lost, if one considers that only part of the nuclear genome was retained by the photosynthetic primary organism during endosymbiosis. Finally, in the bottom right yellow part of the figure, the parallel increase in the contribution of the nuclear genome to plastid function is indicated. Taken together, it covers almost an entire

bacterial genome in land plants, which feature the most complete gene transfer from plastid to nucleus.

Note that the parasitic life of Apicomplexa has led to a consequent decrease in the plastid genome to ~30 genes and the loss of the major photosynthetic genes including RbcL. We do not take this into account in the figure, as parasitism is likely to have biased the process by speeding up gene loss during evolution (van Dooren and Striepen, 2013; Wilson *et al.*, 1996). Such genomes encode mostly remnants of the transcription-translation system and have almost entirely lost the ancestral photosynthetic machinery. The purpose of the persistence of a protein synthesis machinery is unknown.

Legend to Figure Box 1: Evolutionary steps leading to progressive relocalization of the N-terminal processing machinery from the plastid to the nuclear genome along the primary endosymbiosis timescale.

The main Nt processing enzymes from bacteria are listed. The RbcL and RbcS subunits of RuBisCO are used as landmarks and probes of the behavior of a crucial multi-subunit complex requiring complex N-terminal maturation of major subunits.

Panel A: The locations of the corresponding genes along the evolutionary timescale (indicated below) are indicated. An arrow below indicates the corresponding evolutionary timepoint. Loss of *NAA90* has been observed in the plastid of *Cyanophora paradoxa* but not in the other Glaucophyta plastids available to date. This is shown with an asterisk.

Panel B: The type of transfer for each of the genes of the Nt processing machinery shown in Panel A is indicated.

Panel C: The number of open reading frames encoded in the plastid across the evolutionary timescale is indicated for the main groups of plastid-containing organisms resulting from primary endosymbiosis.

Panel D: The number of open reading frames encoded in the nucleus across the evolutionary timescale is indicated.

FIGURE LEGENDS

Figure 1. A schematized tree of life showing the high representation of plastid-containing organisms in eukaryotic biodiversity.

CCTH: Cryptophyta, Centrohelida, Telonemia, and Haptophyta. SAR: Stramenopiles, Alveolata, and Rhizaria. The classification is according to (Lynch *et al.*, 2014). Full or partial genomic data and annotation are available for all the indicated organisms. LECA is Last Eukaryote Common Ancestor. The names of organisms with plastids are colored in light and dark green, brown, and purple. Organisms without plastids are in black. Model organisms frequently mentioned in this report are underlined.

Figure 2. The HUNTER workflow.

The different steps of the HUNTER workflow are displayed (see also main text). The process allows the determination of the N-terminal modification (i.e., Nt-acetylation or free) at the proteome scale

Figure 3. Main modification events at the N-termini of plastid-containing organisms.

Modifications are grouped in three classes (Addition, green; cleavage, blue, alteration, grey) Modifications only retrieved in organelles (i.e., Mitochondria and plastids) are colored in orange.

Figure 4. NAT specificity and the unique amino terminal processing modes of actin.

Panel A: Diverse NATs and their substrate specificity in plastid-containing organisms. NatA, B, H are involved in the processing of actins (panel B). The different NatGs of *A. thaliana* are

mentioned. The number in front of each Nat relates to the percentage of substrate proteins in the proteome (Aksnes *et al.*, 2016). X means any amino acid.

Panel B: Types of actin processing in the cytosol and the NTPM and associated machinery involved. ANAP is N-terminal acetyl-aminopeptidase. **Table 4** provides different examples of the processing types in plastid-containing organisms.

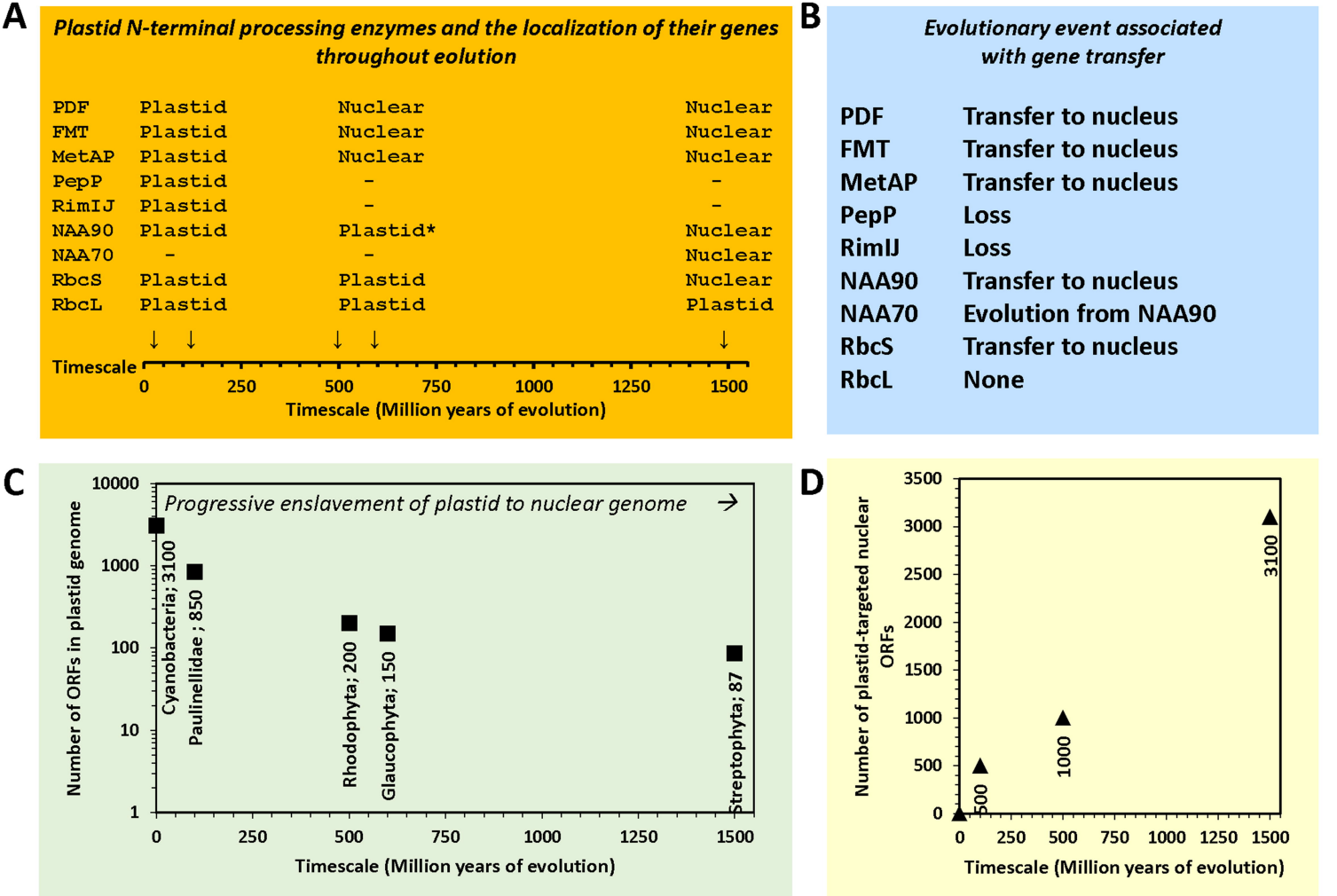

Figure Box 1: Evolutionary steps leading to progressive relocalization of the N-terminal processing machinery from the plastid to the nuclear genome along the primary endosymbiosis timescale.

The main Nt processing enzymes from bacteria are listed. The RbcL and RbcS subunits of RuBisCO are used as landmarks and probes of the behavior of a crucial multi-subunit complex requiring complex N-terminal maturation of major subunits. *Panel A*: The locations of the corresponding genes along the evolutionary timescale (indicated below) are indicated. An arrow below indicates the corresponding evolutionary timepoint. Loss of *NAA90* has been observed in the plastid of *Cyanophora paradoxa* but not in the other Glaucophyta plastids available to date. This is shown with an asterisk. *Panel B*: The type of transfer for each of the genes of the Nt processing machinery shown in Panel A is indicated. *Panel C*: The number of open reading frames encoded in the plastid across the evolutionary timescale is indicated for the main groups of plastid-containing organisms resulting from primary endosymbiosis. *Panel D*: The number of open reading frames encoded in the nucleus across the evolutionary timescale is indicated.

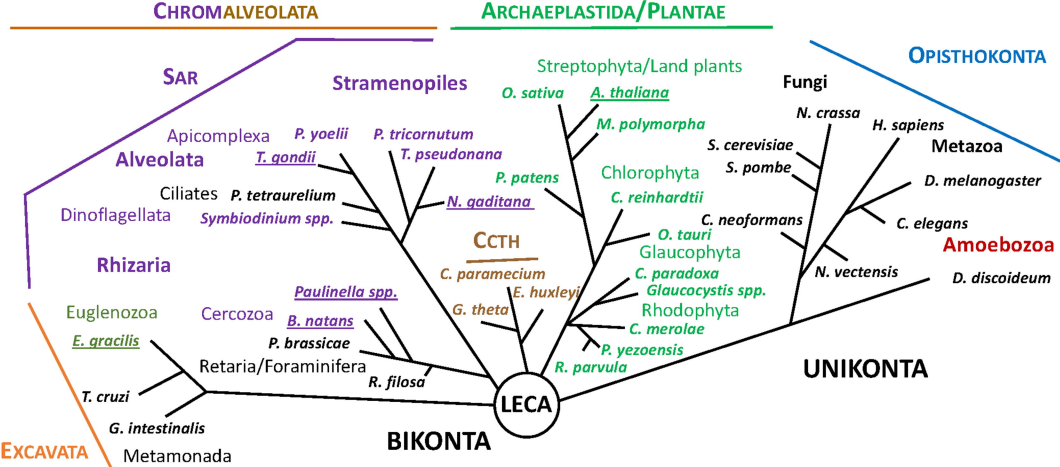

Figure 1. A schematized tree of life showing the high representation of plastid-containing organisms in eukaryotic biodiversity. CCTH: Cryptophyta, Centrohelida, Telonemia, and Haptophyta. SAR: Stramenopiles, Alveolata, and Rhizaria. The classification is according to (Lynch *et al.*, 2014). Full or partial genomic data and annotation are available for all the indicated organisms. LECA is Last Eukaryote Common Ancestor. The names of organisms with plastids are colored in light and dark green, brown, and purple. Organisms without plastids are in black. Model organisms frequently mentioned in this report are underlined.

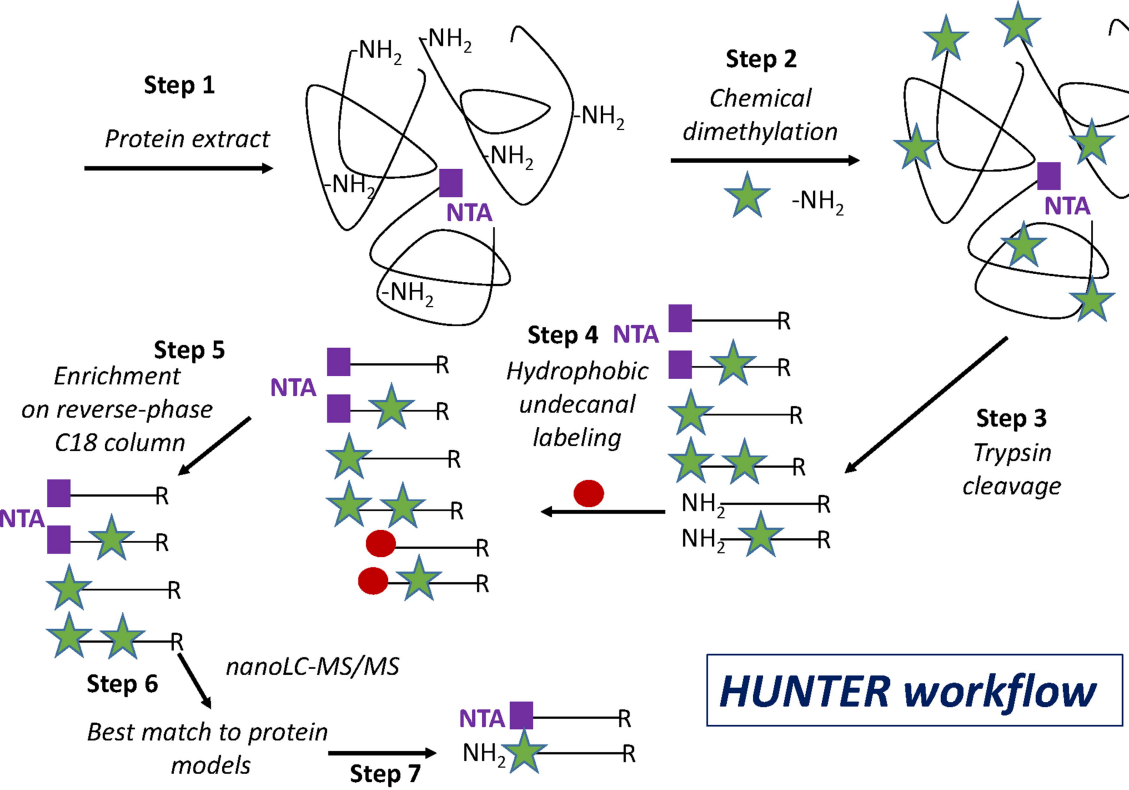

Figure 2. The HUNTER workflow. The different steps of the HUNTER workflow are displayed (see also main text). The process allows the determination of the N-terminal modification (i.e., Nt-acetylation or free) at the proteome scale

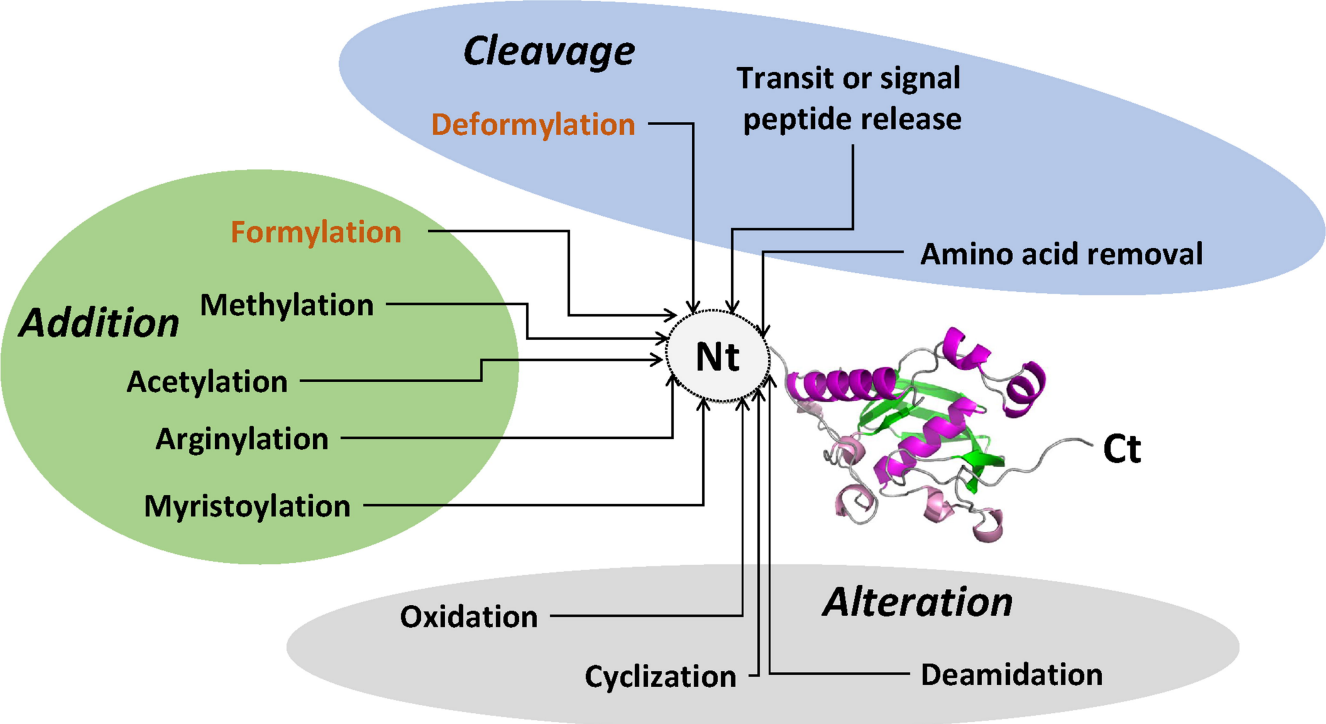

Figure 3. Main modification events at the N-termini of plastid-containing organisms.
Modifications are grouped in three classes (Addition, green; cleavage, blue, alteration, grey) Modifications only retrieved in organelles (i.e., Mitochondria and plastids) are colored in orange.

A

Nuclear-encoded Nats		*Plastid-resident Nats*
NatA: NH_2-ACGSTV	46%	
NatB: NH_2-M(DENQ)	21%	
NatC: NH_2-M(KFIYL)	21%	NatG family - NAA90:GNAT1/2/3 - NAA70:GNAT4/5/6/7/10 NH_2-X
NatE: NH_2-M(AGV)		
NatF: NH_2-M(LV)		
NatH: NH_2-(DE) of actins		
NatD: NH_2-SGRK of histones H2&4		

B

Actin processing (cytosol)

Type Ia&b $M_1(ED)(DE) \xrightarrow{NatB} Ac\text{-}M(ED)$

Type Ib $Ac\text{-}M(ED) \xrightarrow{ANAP} NH_2(DE)_2 \xrightarrow{NatH} Ac\text{-}(DE)_2$

Type IIa&b $M(CA)_2(DE)_3 \xrightarrow{MetAP} NH_2(CA)_2(DE)_3 \xrightarrow{NatA} Ac\text{-}(CA)(DE)_3$

Type IIb $Ac\text{-}(CA)_2(DE)_3 \xrightarrow{ANAP} NH_2(DE)_3 \xrightarrow{NatH} Ac\text{-}(DE)_3$

Figure 4. NAT specificity and the unique amino terminal processing modes of actin. *Panel A*: Diverse NATs and their substrate specificity in plastid-containing organisms. NatA, B, H are involved in the processing of actins (panel B). The different NatGs of *A. thaliana* are mentioned. The number in front of each Nat relates to the percentage of substrate proteins in the proteome (Aksnes *et al.*, 2016). X means any amino acid.
Panel B: Types of actin processing in the cytosol and the NTPM and associated machinery involved. ANAP is N-terminal acetyl-aminopeptidase. **Table 4** provides different examples of the processing types in plastid-containing organisms.