Evolution-driven versatility of N-terminal acetylation in photoautotrophs

Carmela Giglione, Thierry Meinnel

Evolution-driven versatility of N-terminal acetylation in photoautotrophs

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

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

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

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

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

Keywords: N-terminal α-acetyltransferase; lysine ε-acetyltransferase; GNAT; plastid; plant development; stress response

ABSTRACT

N-terminal protein α-acetylation (NTA) is a pervasive protein modification that has recently attracted renewed interest. Early studies on NTA were mostly conducted in yeast and metazoans, providing a detailed portrait of the modification, which was indirectly applied to all eukaryotes. However, new findings originating from photosynthetic organisms have expanded our knowledge of this modification, revealing strong similarities as well as idiosyncratic features. Here, we review the most recent advances on NTA and its dedicated machinery in photosynthetic organisms. We discuss the cytosolic and unique plastid NTA machineries and their critical biological roles in development, stress responses, protein translocation and stability. These new findings suggest that the multitasking plastid and cytosolic machineries evolved to support the specific needs of photoautotrophs.

N-terminal-α-protein acetylation in brief

N-terminal-α-acetylation (NTA) is the closest but less well-known cousin of lysine-ε-acetylation (KA). NTA has attracted significant interest over the last decade thanks to a series of discoveries highlighting its substrates, functions, and modifiers at the molecular, cellular, and organism levels and its impact on many important cellular functions. Like KA, NTA is an ubiquitous modification mainly characterized in yeast, humans, insects, and few archaea [1]. NTA is generally regarded as the enzymatic co- or post-translational addition of an acetate moiety onto protein N-α-amino groups. Although NTA is found in all organisms, the percentage of proteins undergoing this modification within a proteome increases with organism complexity, affecting 1-2% of proteins in bacteria [2-4], 10-45% in archaea, 60% in fungi, and >80% of the cytosolic proteome in multicellular eukaryotes [1, 5].

Increasing acetylation from photosynthetic bacteria to plants

Cytosolic NTA in photosynthetic eukaryotes

The N-terminal acetylomes of cyanobacteria or green sulfur bacteria have revealed only a few N-terminal protein acetylation events in photosynthetic bacteria (**Table 1**). By contrast, the characterization of thousands of N-termini in green algae and land plants (i.e., *Arabidopsis thaliana* or *Populus trichocarpa*) has revealed a cytosolic N-terminal acetylome resembling - both in terms of extent (~80%) [6] and specificity - the animal acetylome [6-18]. In the diatom *Thalassiosira pseudonana*, a unicellular photosynthetic eukaryote [19], the N-terminal acetylome is slightly lower (~70%).

Organellar NTA

Although NTA has yet to be detected in mitochondria [20], both co- and post-translational NTA occur in plastids (Key **Figure 1**). NTA impacts 20-30% of all plastid proteins [9, 21-23] (**Table 1**). In land plants - where <100 proteins are encoded by the plastid genome

and a few dozen have had their N-termini characterized - a significant percentage of plastid-encoded proteins were N-terminal acetylated (NTAed) (**Table 1**).

Extensive N-terminal data are now available describing the major soluble and membrane complexes of plastids in various organisms (**Table 1**; [18]). NTA is widespread in major plastid complex subunits but, remarkably, the core translational/transcriptional plastid machinery is poorly acetylated. This is also true for some complexes including PSI, NDH, or cytochrome b_6f, unlike PSII and ATPase, where most major subunits are NTAed. High NTA levels are similarly observed in the most abundant proteins, namely soluble ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) complex or insoluble light-harvesting complex (LHC). These high NTA ratios in plastid protein complexes appear to be unrelated to the location of protein synthesis, as exemplified by RuBisCO, where each subunit is usually synthesized in the plastid or nuclear genome. Therefore, high levels of NTAed subunits of some plastid complexes correlate with oxygenic photosynthesis, which features PSII, RuBisCO, and the dedicated chlorophyll-binding proteins of LHC. NTA could therefore be related to oxidative damage and provide an additional mechanism by which plastids can manage oxygen production whilst preserving redox homeostasis.

A main feature of plastid NTA is that acetylation yields (i.e. percentage of NTA of one single proteoform) are often low [24, 25]. However, this NTA yield might be dependent on physiological growth conditions including stress, light, and carbon source availability. Full NTA has only been observed for a few proteins including PSII components such as PsbD and PsbS, LHC components, amino acid or lipid synthesis enzymes, redox, and a set of plastid importins (TIC20/55/62, IMB1, SecA [25]). This small set of fully NTAed plastid proteins is an incomplete characterization, since only a reduced set of plastid proteins have been analyzed.

Current state-of-the-art: the plant NAT machinery

General features

NTA is catalyzed by N-α-acetyltransferases (NATs; subunits designated NAAs or GNATs), which via at least one catalytic subunit transfer the acetyl group from acetyl-CoA to the free α-amino groups of protein N-termini. Several cytosolic/membrane NATs have been identified, each showing distinct substrate specificity [1]. In NatA/B/C/E, the corresponding catalytic subunit associates with one or more auxiliary subunits to improve catalytic performance, specificity, and/or ribosome binding [1]. All NAAs belong to the very large general control non-repressible 5 (GCN5)-related N-acetyltransferase (GNAT) superfamily and share a common 3D core architecture [26-28]. Recent work has provided a detailed picture of the NAT machinery in plants. **Table 2** compares the protein entries for *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana*, and *Oryza sativa*, which shows that monocots and dicots display a similar machinery and the same NATs to metazoans except for NAA80, an actin-specific NAT [1]. Analysis of the machinery of green plants shows strong conservation in the phylum (**Box 1**).

The NatA/NatE core complex

A. thaliana NatA was the first NAT complex characterized in a photosynthetic organism [13, 15]. The core of this complex consists of a catalytically active subunit (AtNAA10) and an auxiliary subunit AtNAA15 (**Table 2**). Yeast and animal NAA15s serve as ribosome-binding adaptors and modify NAA10's intrinsic activity [29]. Human NAA10 also exists as a monomer and, at least *in vitro*, acetylates the α-amino group of proteins starting with acidic residues (Asp/Glu) that can be generated post-translationally [30]. However, human and yeast NAA10 do not seem to have intrinsic classical NatA-type activity [29]. In complex with the ribosome-binding partner NAA15, yeast and human NAA10 display different substrate specificity, acetylating nascent peptide chains from which the initial methionine (iMet) has been removed

by methionine aminopeptidases (MetAPs) during a co-translational process called N-terminal methionine excision (NME) [31] (**Figure 1**). Due to the substrate specificity of MetAPs, NatA substrates start with residues featuring non-bulky side chains [32]. Like in humans, NAA10 and NAA15 are mandatory for NatA activity *in vivo* in *A. thaliana*, affecting 50% of soluble proteins in leaves [13]. AtNAA15 is anchored to ribosomes [13, 15]. Unlike yeast NatA [33], AtNAA10-AtNAA15 association is unnecessary for typical AtNatA specificity, preferentially affecting proteins undergoing NME.

In humans and a few fungi, the NatA complex also involves the chaperone Huntingtin-interacting protein K (HYPK) [34, 35] (**Table 2**). In yeast and metazoans, NatA forms complexes with another subunit, NAA50 (**Table 2**) [36, 37], defined as NatE [38] and recently identified in *A. thaliana* [39-41]. AtNAA50 is catalytically active with the cognate NAA50 Met-Leu-Gly-Pro peptide, similar to human NAA50. Unlike yeast and fly NAA50 [28, 36, 42], AtNAA50 knockout does not affect the NTA of canonical NatA substrates [39]. It was therefore proposed that (i) the absence of human NAA50 might promote HYPK binding to NatA, which inhibits NatA activity [34, 35], and/or (ii) human NAA50 might contribute to NatA's ribosomal association in multicellular eukaryotes [35, 43, 44], although an absence of effects on *A. thaliana* NatA substrates disfavors the second hypothesis. In addition to its NTA activity, AtNAA50 was shown *in vitro* to auto-acetylate the ε-amino group of internal lysines (KA activity), as for HsNAA50 [39]. This dual activity of AtNAA50 seems to fit with its nucleocytoplasmic and endoplasmic reticulum localization (see below and [39]). It was noticed that AtNAA50 does not co-immunoprecipitate with AtNatA [40].

The NatB complex

In metazoans and yeast, the NatB complex is composed of the catalytic subunit NAA20 and the auxiliary subunit NAA25 [1] (**Table 2**), specifically acetylating the N-termini of Met-Glu/Asp/Asn/Gln-starting proteins [45]. The first evidence of photosynthetic NatB activity on

a Met-Glu substrate was provided by *in vitro* analysis of recombinant AtNAA20 and by silencing of *NAA25* in tobacco [15]. Plant NatB recapitulates the specificity of NatB in other organisms [14]. NatB is predicted to contribute 20% of *A. thaliana*'s cytosolic NTAed proteome.

Despite this conserved substrate specificity, only HsNAA20, but not ScNAA20, rescued NAA20-deficient *A. thaliana naa20-1* lines, suggesting functional assembly with endogenous AtNAA25 [14].

The NatC complex

The NatC complex consists of the catalytic subunit NAA30 and the auxiliary subunits NAA35 and NAA38 (**Table 2**). NatC acetylates iMet-starting proteins, favoring hydrophobic and basic amino acids at position 2 in both yeast and humans [46, 47]. However, such proteins are not systematically NTAed, and other substrate specificity determinants are likely to be involved downstream of position 2 [48, 49].

There is currently no detailed analysis of plant NatC substrate specificity. Nevertheless, functional complementation of the loss-of-function *mak3* yeast mutant with a putative AtNAA30 provided evidence that this subunit has NatC activity [50]. NatC activity was independent of the other two yeast auxiliary subunits, suggesting that AtNAA30 does not require NatC complex formation for enzymatic activity. However, interaction between AtNAA30 and AtNAA35, but not two potential NAA38s, was shown [50]. It is likely that the putative NAA38 subunits were not properly identified in these experiments (see alternative genes in **Table 2**).

NatF/Naa60

NatF is a monomer formed from the sole catalytic subunit NAA60, so far only identified in multicellular organisms including plants. Five orthologs have been identified in humans, mainly differing at the N- or C-terminus. Similarly, the three identified *A. thaliana* orthologs

differ at their C-terminus. One HsNAA60 isoform is anchored to Golgi membranes [51], whereas an AtNAA60 isoform showed C-terminal-dependent plasma membrane localization. It will be interesting to establish how the C-termini of the other two orthologs influence membrane localization.

AtNAA60 showed greater KA activity *in vitro* than HsNAA60. 3D structures of AtNAA60 in complex with either AcetylCoA or the bisubstrate analogue CoA-Ac-MetValAsnAlaLeu [52] showed the characteristic folded GNAT architecture with active site residue arrangements similar to the HsNAA60 structures, including a hydrophobic binding pocket perfectly matching the high selectivity for iMet. Substitution of conserved residues of the active site or the substrate-binding pocket of AtNAA60 demonstrated mechanistic conservation with human NAA50 and NAA60 [52]. Nevertheless, AtNAA60 is monomeric in solution, whilst HsNAA60 forms a homodimer with acetyl-CoA [53, 54], presumably due to the difference in the β6–β7 loop region sequence. Finally, N-terminus acetylome profiling of AtNAA60 KO mutants revealed one plasmodesmata-localized protein, At5g03660, as a specific AtNAA60 substrate [52].

A plastid-targeted family of active GNATs featuring dual KA and NTA activity

The high frequency of plastid NTA observed in land plants suggests that there is at least one plastid NAT with large substrate specificity or several isoforms with narrower specificity, as in the cytosol [9]. Several putative candidates have been proposed in *A. thaliana* and other land plants, collectively called NatG (see above and [5, 9, 25]). AtGNAT4 (initially called NAA70, now one of the members of this family) was the first plastid NAT to be characterized [55]. Characterization of ten putative plastid *A. thaliana* GNAT proteins, including AtGNAT4 [25], revealed that eight were plastid localized and all exhibited the typical GNAT topology but with several divergences from the cytosolic NAT enzymes [55]. GAP assays [25] revealed strong NTA activity for six of the eight plastid-associated GNATs (**Box 2**). NTA activity of

GNAT1 and GNAT3 was inefficient, most likely because they (i) act on a restricted number of plastid substrates or (ii) require plant-specific accessory proteins to improve their activity. All GNAT candidates had a broad range of substrates (**Box 2**). Unexpectedly, all plastid-associated GNATs displayed specific KA activity, although again this was weak for GNAT1 and GNAT3. The dual activity of AtGNAT2, also known as NSI (nuclear shuttle interacting), was demonstrated by AtGNAT2/NSI knockout in two independent backgrounds (*nsi-1* and *nsi-2*) [25, 56, 57], with global quantitative NTA and KA analyses showing that NTA and KA activity were affected in AtGNAT2-defective plants. The most sensitive KA and NTA targets in AtGNAT2-defective plant lines were different, suggesting a different acetylation recognition mode for each of the two activities. The N-termini NTAs of six major plastid proteins were affected in the AtGNAT2-defective plant lines (see Table 1 in [25]). Interestingly, NTA of different N-termini of the same protein were affected in AtGNAT2-defective plant lines, suggesting long distance contacts between AtGNAT2 and their substrates, perhaps explaining the specific selectivity of AtGNAT2 for some substrates.

GNAT2 was initially described in rice and in *A. thaliana* as a serotonin acetyltransferase (SNAT1; [58-60]). However, it is a surprisingly weak enzyme, 2-3 orders of magnitude less active than the mammalian homolog [59]. A second SNAT2 was also identified in rice based on its close similarity with OsSNAT1 [59], corresponding to AtGNAT1. It is likely that GNATs of the former SNAT family also display weak activity on small substrates such as serotonin, highlighting their relaxed specificity. Finally, the plastid-encoded SNAT from the red alga *Pyropia yezoensis* displays close homology with AtGNAT2 [47] and is most likely a GNAT.

Physiological relevance of NATs in phototrophic organisms

The multiple impacts of impaired plant cytosolic machineries

NTA regulates diverse molecular functions in yeast and metazoans [1]. In *A. thaliana*, *NAA10* and *NAA15* are classified as embryo defective genes [61]. Indeed, characterization of

A. thaliana T-DNA insertion mutant *NAA10* and *NAA15* lines confirmed that loss of both subunits arrested embryos at the dermatogen to early globular stage and that AtNatA is required for the asymmetric division of the hypophysis, root meristem formation, and proper suspensor development [13, 62, 63]. Additionally, artificial miRNA (*ami*) AtNAA10 and AtNAA15 lines displayed severe developmental defects and growth retardation compared to wild-type *A. thaliana*, showing that both subunits are necessary for NatA activity. *amiNAA10* and *amiNAA15* were also highly drought tolerant, presumably due to altered root morphology and reduced stomatal apertures (**Figure 2**). Given that many drought signaling-associated genes, including those associated with the phytohormone abscisic acid (ABA), were overexpressed in *amiNAA10* and *amiNAA15* lines, the extreme drought stress tolerance observed with reduced NatA activity was ascribed to alterations in ABA perception. Consistent with that hypothesis, exogenous ABA or drought stress in wild-type plants rapidly depleted NatA transcripts, finally inducing several adaptive responses that improved plant drought tolerance and survival. No N-terminal acetylome of drought-stressed plants has yet been reported. However, the use of a fluorescent probe, which specifically reacts at neutral pH with the N-terminal amino groups of polypeptides, revealed an increase of free N-termini in response to drought and ABA [13]. Although definitive proof is still missing, these data suggest that the proportion of free N-termini might also increase in these conditions. Therefore, it was proposed that NTA by NatA is a hormone-controlled dynamic process that regulates cellular stress responses to drought in higher plants. How this NTA dynamicity is controlled is unknown. Since the proteasome machinery was a major upregulated pathway in *amiNAA10* and *amiNAA15*, AtNatA might control cellular proteostasis.

Loss of *A. thaliana* NAA50 function produced dwarf plants with altered root morphology, infertility, and induced stress signaling responses comparable to the effects of reduced NatA in plants [15, 39, 40]. Accordingly, reduction of *A. thaliana* NAA50 arrested

stem and root growth and senescence and altered developmental process genes and proteins mediating stress responses [39, 41]. Additionally, reduced AtNAA50 constitutively activated endoplasmic reticulum (ER) stress signaling, suggesting that AtNAA50 might also play a role in regulating ER stress in plants (e.g., preventing protein misfolding and/or aggregation). Interestingly, it has been proposed that indirect AtNAA50-mediated regulation of ER stress is the primary mechanism responsible for the observed increase in stress signaling in AtNAA50 knockout and knockdown plants (**Figure 2**) [41]. For instance, if NTA is required to prevent induction of ER stress, then the enhanced resistance to different abiotic stresses may be an indirect result of changes to ER stress signaling. Although interactions between AtNAA50 and the *A. thaliana* protein kinase EDR1 (enhanced disease resistance 1), a negative regulator of cell death during biotic and abiotic stress responses, have been demonstrated, and loss of EDR1 leads to enhanced ER stress sensitivity, how EDR1 in association with AtNAA50 regulates ER stress is unknown [41].

Matsuo et al. [64] first reported photosynthetic NatB in 2012, identifying the catalytic NatB subunit in *C. reinhardtii* as responsible for the circadian rhythm of rhythm of chloroplast 97 (roc97) mutants. NatB might therefore be essential for maintenance of a robust circadian clock in *C. reinhardtii*. In *A. thaliana*, AtNatB has been implicated in abiotic stress responses in addition to playing key roles in vegetative and reproductive developmental processes [14, 65]. Indeed, downregulation of both AtNAA20 or AtNAA25 subunits impacted growth and resistance to osmotic and high-salt stresses [14]. Finally, a genetic interaction has been proposed between AtNAA25 and ARGONAUTE10, but the nature of these interactions has not been established [65].

It has also been reported that NatB, by antagonizing NatA, controls the stability of the immune receptor SUPPRESSOR OF NPR1, CONSTITUTIVE 1 (SNC1) [15]. Indeed, targeted genetic and biochemical analyses of SNC1 revealed that NatA-mediated acetylation of the N-

terminal variant MetMetAsp-SNC1 strongly destabilizes the protein, whereas NatB-mediated acetylation of the N-terminal variant MetAsp-SNC1 stabilizes the immune receptor (**Figure 2**) [15]. Therefore, NTA may have opposing effects on a single protein depending on the sequence context, and the complex regulation afforded by different NAT complexes might provide flexibility to a specific target protein and faster attainment of appropriate protein concentrations under dynamic cellular conditions [15, 66-70]. In agreement, NatB-mediated NTA stabilized SIGMA FACTOR-BINDING PROTEIN 1 (SIB1) [71], a transcription co-regulator targeted to both the nucleus and chloroplasts. SIB1 is implicated in reactive oxygen species- and salicylic acid-mediated stress responses and represses the expression of photosynthesis-associated genes [72, 73].

An *A. thaliana* mutant line defective in the NatC component NAA30 was originally identified based on its photosynthetic phenotype of reduced photosystem II performance [50]. This mutant also exhibited smaller rosettes, paler leaves, and fewer thylakoid complexes (**Figure 2**). To explain how cytosolic NatC-dependent NTA affected photosynthesis, NTA was suggested to contribute to the stability and/or import competence of organellar precursor proteins. Supporting this hypothesis, in Toc-deficient plant lines, NTAed plastid precursor proteins accumulated in the cytosol [22, 74], and most of these accumulated nuclear-encoded plastid-imported proteins were found to be NTAed in the cytosol at the level of the N-termini of uncleaved transit peptides. Therefore, nuclear-encoded plastid-imported proteins might undergo NTA twice, first in the cytosol during translation and second on neo-N-termini in the plastid after transit peptide cleavage (**Figure 1**). Consistent with the need for transit peptide protein modifications before cleavage, cytoplasmic phosphorylation of chloroplast transit peptides has been shown to be necessary for efficient protein import into the chloroplast [75]. In addition, NTA of the non-cleavable mitochondria-targeting signal of rat chaperonin 10 facilitated mitochondrial import by increasing N-terminal helix stability [76].

Depletion of *A. thaliana* NAA60 does not cause any visible phenotype under standard growth conditions; however, the mutant was hypersensitive to salinity (**Figure 2**), providing evidence that membrane protein NTA is important during stress [52]. However, how NAA60 contributes to salinity resistance is unknown.

Plastid NTA impairment and its impact on photosynthesis

Our understanding of the physiological relevance of plastid GNATs remains in its infancy and is made difficult by the partially overlapping and dual NTA and KA activities of all members. Loss of *A. thaliana* NSI/GNAT2 caused a small reduction in plant growth under control conditions [56, 57, 77] but severe phenotypes in fluctuating light conditions (**Figure 2**); NSI/GNAT2 deletion resulted in *A. thaliana* being unable to perform state transitions. Although LHCII protein phosphorylation/dephosphorylation usually mediates state transitions [78], LHCII phosphorylation was not impaired in NSI/GNAT2 mutants [56, 57]. KA and NTA acetylomes revealed that NSI/GNAT2 ablation decreased both plastid KA and NTA [25, 56]. However, GNAT2 inactivation differentially affected its KA and NTA targets with several photosynthetic apparatus components, including the LCHII subunits only affected in their KA. This led to the suggestion that GNAT2/NSI KA of LHCII is likely to be the posttranslational modification required for state transition in *A. thaliana* [56]. The function of GNAT2/NSI-guided plastid NTA is unknown. Interestingly, *A. thaliana GNAT2/NSI* knockout mutants are more susceptible to pathogen infection due to its SNAT activity [77]. Although this activity was only demonstrated for GNAT2/NSI orthologues in rice and not in *A. thaliana*, decreases in melatonin and salicylic acid levels were observed in *A. thaliana GNAT/NSI* mutant lines.

Concluding remarks and future perspectives

Major efforts are now underway to characterize all the major NAT complexes and NTA in plants. The available data have already revealed a much more sophisticated NAT machinery in plants than expected when compared to humans and fungi. Although plants lack the actin-

specific NAA80, they all have dedicated NATs – NAA70 and NAA90 - involved in plastid protein NTA and KA. Within this family, the *Pyropia* GNAT is a plastid remnant evidencing the endosymbiotic origin of the family. It is therefore unsurprising that cyanobacteria have specific AtGNAT2 homologs in addition to the usual bacterial-origin NATs, since photosynthetic bacteria were fully expected to display a dedicated NAT machinery, most likely for NTA of major PSII subunits (**Table 1**).

Plastid NTA is now a major topic in the NAT research field. However, several issues still need to be addressed. There is a need for dedicated tools to: (i) characterize the specific acetylome for each GNAT; (ii) uncouple the KA from the NTA activity of each GNAT; and (iii) discriminate the impact of NTA activity on plastid function from overall cellular functions. In this respect, it is worth remembering that inhibition of protein N-terminal deformylation - an organellar co-translational process that can be specifically blocked in plastids - in chlorophyta and spermatophytes decreases the accumulation of only four plastid proteins, all belonging to the PSII complex, namely the core protein components PsbA/B/C/D [79-82]. This decreased stability was hypothesized to be caused by absent NME [81] or the presence of formyl [83], but the absence of NTA has not yet been excluded even though normal NTA is blocked in the four destabilized subunits due to deformylation inhibition (**Figure 1**). In addition, the impact of NTA on directly adjacent Thr phosphorylation has yet to be considered.

Partial plastid NTA [24] might occur for a number of reasons. First, the involved machinery might be poorly efficient due to its post-translational nature in the absence of known interactions with plastoribosomes. Whether interactions with the ribosome or the import machinery are via auxiliary subunits or direct will need to be investigated. An absence of channeling between the substrate and the catalyst at the ribosome, as occurs in the cytosol for NatA/B/C, could significantly contribute to poor NTA efficiency. Second, plastid NTA might have initially involved only a few proteins, as in cyanobacteria. The progressive migration of

many genes to the nucleus during the course of plastid evolution and the use of an import machinery leading to transit peptides will have opened up many new, fortuitous targets. As a result, their NTA would be an unnecessary decoration. If this hypothesis is true, then only fully NTAed proteins would really matter physiologically, but this also means that this mechanism would involve significant (and unexpected) AcCoA wastage. Furthermore, LHC components arising from the nuclear genome appear to display high NTA yields, making this hypothesis unlikely. Third, plastid protein acetylation might contribute to protein stability/turnover. Supporting this hypothesis, the chloroplast-encoded AtpE subunit is partially NTAed in several land plants [84, 85]. Upon drought stress in watermelon, the non-NTAed form of the AtpE subunit dropped from 72% to 11%, whereas the NTAed form remained constant [84], the degradation of the former proposed to be mediated by a stress-induced metalloprotease. Accordingly, short-lived stromal proteins in *C. reinhardtii* were also non-NTAed, which suggests that NTA promotes plastid protein stability [86]. Alternatively, we propose that plastid NTA-induced stability is influenced by a combination of modifications. Remember that PSII turnover often involves vicinal phosphorylation and NTA, as in the case of PsbD in response to heat or high light stress. Finally, NTAed proteins might be involved in the protein-protein interactions required in the complexes, whereas the unacetylated isoforms could constitute a reservoir, progressive acetylation of which would allow post-translational replacement of the oxygen-damaged proteins. Such recycling without protein neo-synthesis would be well suited for swift adaptation of plants from low to high light conditions (sunshine/cloudy). Accordingly, NTA of the PSII and LHCII subunits was recently shown to be crucial for stacking in the thylakoid membrane [87].

The discovery of the association between cytosolic photosynthetic NATs and many stress-related pathways and retrograde signaling with the plastid machinery opens up new avenues for studies exploring the molecular mechanisms linking NAT-dependent NTA to plant

stress responses and the impact of NTA on communication between different cellular compartments.

ACKNOWLEDGEMENTS

This research of the team is funded by the French Agence Nationale de la Recherche (ANR-13-BSV6-0004, ANR-17-CAPS-0001-01) 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).

SUPPLEMENTAL INFORMATION

Supplemental information associated with this article can be found at doi:XXXXXXX'

REFERENCES

1. Aksnes, H. et al. (2019) Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. Mol Cell 73 (6), 1097-1114.

2. Bienvenut, W.V. et al. (2015) Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition. Proteomics 15 (14), 2503-18.

3. Schmidt, A. et al. (2016) The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol 34 (1), 104-10.

4. Ouidir, T. et al. (2015) Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14. J Proteomics 114C, 214-225.

5. Breiman, A. et al. (2016) The intriguing realm of protein biogenesis: Facing the green co-translational protein maturation networks. Biochim Biophys Acta 1864 (5), 531-550.

6. Liu, C.C. et al. (2013) Identification and analysis of the acetylated status of poplar proteins reveals analogous N-terminal protein processing mechanisms with other eukaryotes. PLoS One 8 (3), e58681.

7. Martinez, A. et al. (2008) Extent of N-terminal modifications in cytosolic proteins from eukaryotes. Proteomics 8 (14), 2809-2831.

8. Baerenfaller, K. et al. (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science 320 (5878), 938-41.

9. Bienvenut, W.V. et al. (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. Mol Cell Proteomics 11 (6), M111 015131.

10. Zhang, H. et al. (2018) N-terminomics reveals control of Arabidopsis seed storage proteins and proteases by the Arg/N-end rule pathway. New Phytol 218 (3), 1106-1126.

11. Zhang, H. et al. (2015) Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots. Proteomics 15 (14), 2447-2457.

12. Venne, A.S. et al. (2015) An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in Arabidopsis thaliana. Proteomics 15 (14), 2458-69.

13. Linster, E. et al. (2015) Downregulation of N-terminal acetylation triggers ABA-mediated drought responses in Arabidopsis. Nat Commun 6, 7640.

14. Huber, M. et al. (2020) NatB-mediated N-terminal acetylation affects growth and abiotic stress responses. Plant Physiol 182, 792-806.

15. Xu, F. et al. (2015) Two N-terminal acetyltransferases antagonistically regulate the stability of a nod-like receptor in Arabidopsis. Plant Cell 27 (5), 1547-62.

16. Sun, Q. et al. (2009) PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Res 37 (Database issue), D969-74.

17. Soh, W.T. et al. (2020) ExteNDing proteome coverage with legumain as a highly specific digestion protease. Anal Chem 92 (4), 2961-2971.

18. Millar, A.H. et al. (2019) The scope, functions, and dynamics of posttranslational protein modifications. Annu Rev Plant Biol 70, 119-151.

19. Huesgen, P.F. et al. (2013) Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. PLoS One 8 (9), e74483.

20. Giglione, C. and Meinnel, T. (2001) Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. Trends Plant Sci. 6 (12), 566-572.

21. Zybailov, B. et al. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One 3 (4), e1994.

22. Bischof, S. et al. (2011) Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. Plant Cell 23 (11), 3911-28.

23. Rowland, E. et al. (2015) The Arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. Plant Physiol 169 (3), 1881-96.

24. Bouchnak, I. and van Wijk, K.J. (2019) N-Degron Pathways in Plastids. Trends Plant Sci 24 (10), 917-926.

25. Bienvenut, W.V. et al. (2020) Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. Mol Syst Biol 16 (7), e9464.

26. Vetting, M.W. et al. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. Arch Biochem Biophys 433 (1), 212-26.

421 27. Salah Ud-Din, A.I. et al. (2016) Structure and functional diversity of GCN5-related N-
422 acetyltransferases (GNAT). Int J Mol Sci 17 (7), E1018.

423 28. Rathore, O.S. et al. (2016) Absence of N-terminal acetyltransferase diversification during
424 evolution of eukaryotic organisms. Sci Rep 6, 21304.

425 29. Ree, R. et al. (2018) Spotlight on protein N-terminal acetylation. Exp Mol Med 50 (7), 90.

426 30. Van Damme, P. et al. (2011) Proteome-derived peptide libraries allow detailed analysis of
427 the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-
428 translational actin N(alpha)-acetyltransferase. Mol Cell Proteomics 10 (5), M110 004580.

429 31. Giglione, C. et al. (2015) N-terminal protein modifications: Bringing back into play the
430 ribosome. Biochimie 114, 134-146.

431 32. Arnesen, T. et al. (2009) Proteomics analyses reveal the evolutionary conservation and
432 divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci U S A
433 106 (20), 8157-8162.

434 33. Liszczak, G. et al. (2013) Molecular basis for N-terminal acetylation by the heterodimeric
435 NatA complex. Nat Struct Mol Biol 20 (9), 1098-105.

436 34. Weyer, F.A. et al. (2017) Structural basis of HypK regulating N-terminal acetylation by the
437 NatA complex. Nat Commun 8, 15726.

438 35. Deng, S. et al. (2020) Molecular basis for N-terminal acetylation by human NatE and its
439 modulation by HYPK. Nat Commun 11 (1), 818.

440 36. Gautschi, M. et al. (2003) The yeast N(alpha)-acetyltransferase NatA is quantitatively
441 anchored to the ribosome and interacts with nascent polypeptides. Mol Cell Biol 23 (20), 7403-
442 14.

443 37. Evjenth, R. et al. (2009) Human Naa50p (Nat5/San) displays both protein Nα- and Nϵ-
444 acetyltransferase activity. J Biol Chem 284 (45), 31122-31129.

445 38. Aksnes, H. et al. (2016) First things first: vital protein marks by N-terminal
446 acetyltransferases. Trends Biochem Sci 41 (9), 746-60.

447 39. Armbruster, L. et al. (2020) NAA50 is an enzymatically active Nα-acetyltransferase that is
448 crucial for development and regulation of stress responses. Plant Physiol 183(4), 1502-16.

449 40. Feng, J. et al. (2020) The N-Terminal acetyltransferase Naa50 regulates Arabidopsis growth
450 and osmotic stress response. Plant Cell Physiol 61 (9), 1565-1575.

451 41. Neubauer, M. and Innes, R.W. (2020) Loss of the acetyltransferase NAA50 Induces ER
452 stress and immune responses and suppresses growth. Plant Physiol 183 (4), 1838-1854.

453 42. Van Damme, P. et al. (2015) N-terminal acetylome analysis reveals the specificity of Naa50
454 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and
455 methionine aminopeptidases. Proteomics 15 (14), 2436-46.

456 43. Knorr, A.G. et al. (2019) Ribosome–NatA architecture reveals that rRNA expansion
457 segments coordinate N-terminal acetylation. Nat Struct Mol Biol 26 (1), 35-39.

458 44. Deng, S. et al. (2019) Structure and mechanism of acetylation by the N-terminal dual
459 enzyme NatA/Naa50 complex. Structure 27 (7), 1057-1070.e4.

460 45. Van Damme, P. et al. (2012) N-terminal acetylome analyses and functional insights of the
461 N-terminal acetyltransferase NatB. Proc Natl Acad Sci U S A 109 (31), 12449-54.

462 46. Polevoda, B. et al. (1999) Identification and specificities of N-terminal acetyltransferases
463 from *Saccharomyces cerevisiae*. EMBO J 18 (21), 6155-6168.

464 47. Van Damme, P. et al. (2016) A role for human N-alpha Acetyltransferase 30 (Naa30) in
465 maintaining mitochondrial integrity. Mol Cell Proteomics 15 (11), 3361-3372.

466 48. Kimura, Y. et al. (2003) N-Terminal modifications of the 19S regulatory particle subunits
467 of the yeast proteasome. Arch Biochem Biophys 409 (2), 341-348.

468 49. Starheim, K.K. et al. (2009) Knockdown of human N alpha-terminal acetyltransferase
469 complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. Mol Cell
470 Biol 29 (13), 3569-81.

471 50. Pesaresi, P. et al. (2003) Cytoplasmic N-terminal protein acetylation is required for efficient
472 photosynthesis in Arabidopsis. Plant Cell 15 (8), 1817-32.

473 51. Aksnes, H. et al. (2017) Molecular determinants of the N-terminal acetyltransferase Naa60
474 anchoring to the Golgi membrane. J Biol Chem 292 (16), 6821-6837.

475 52. Linster, E. et al. (2020) The Arabidopsis Nα-acetyltransferase NAA60 locates to the plasma
476 membrane and is vital for the high salt stress response. New Phytol 228 (2), 554-569.

477 53. Chen, J.Y. et al. (2016) Structure and function of human Naa60 (NatF), a Golgi-localized
478 bi-functional acetyltransferase. Sci Rep 6, 31425.

479 54. Støve, S.I. et al. (2016) Crystal structure of the Golgi-associated human Nα-
480 acetyltransferase 60 reveals the molecular determinants for substrate-specific acetylation.
481 Structure 24 (7), 1044-56.

482 55. Dinh, T.V. et al. (2015) Molecular identification and functional characterization of the first
483 Nalpha-acetyltransferase in plastids by global acetylome profiling. Proteomics 15 (14), 2426-
484 35.

485 56. Koskela, M.M. et al. (2018) Chloroplast acetyltransferase NSI is required for state
486 transitions in *Arabidopsis thaliana*. Plant Cell 30 (8), 1695-1709.

487 57. Koskela, M.M. et al. (2020) Comparative analysis of thylakoid protein complexes in state
488 transition mutants nsi and stn7: focus on PSI and LHCII. Photosynth Res 145 (1), 15-30.

489 58. Kang, K. et al. (2013) Molecular cloning of rice serotonin N-acetyltransferase, the
490 penultimate gene in plant melatonin biosynthesis. J Pineal Res 55 (1), 7-13.

491 59. Byeon, Y. et al. (2016) Cloning and characterization of the serotonin N-acetyltransferase-2
492 gene (SNAT2) in rice (Oryza sativa). J Pineal Res 61 (2), 198-207.

493 60. Lee, H.Y. et al. (2014) Cloning of Arabidopsis serotonin N-acetyltransferase and its role
494 with caffeic acid O-methyltransferase in the biosynthesis of melatonin in vitro despite their
495 different subcellular localizations. J Pineal Res 57 (4), 418-426.

496 61. Devic, M. (2008) The importance of being essential: EMBRYO-DEFECTIVE genes in
497 Arabidopsis. C R Biol 331 (10), 726-36.

498 62. Feng, J. et al. (2016) Protein N-terminal acetylation is required for embryogenesis in
499 Arabidopsis. J Exp Bot 67 (15), 4779-89.

500 63. Feng, J. and Ma, L. (2016) NatA is required for suspensor development in Arabidopsis.
501 Plant Signal Behav 11 (10), e1231293.

502 64. Matsuo, T. et al. (2012) N-terminal acetyltransferase 3 gene is essential for robust circadian
503 rhythm of bioluminescence reporter in Chlamydomonas reinhardtii. Biochem Biophys Res
504 Commun 418 (2), 342-346.

505 65. Ferrandez-Ayela, A. et al. (2013) Mutation of an Arabidopsis NatB N-alpha-terminal
506 acetylation complex component causes pleiotropic developmental defects. PLoS One 8 (11),
507 e80697.

508 66. Kapos, P. et al. (2015) N-terminal modifications contribute to flowering time and immune
509 response regulations. Plant Signal Behav 10 (10), e1073874.

510 67. Gibbs, D.J. (2015) Emerging functions for N-terminal protein acetylation in plants. Trends
511 Plant Sci 20 (10), 599-601.

512 68. Linster, E. and Wirtz, M. (2018) N-terminal acetylation: an essential protein modification
513 emerges as an important regulator of stress responses. J Exp Bot 69 (19), 4555-4568.

514 69. Gibbs, D.J. et al. (2016) From start to finish: amino-terminal protein modifications as
515 degradation signals in plants. New Phytol 211 (4), 1188-94.

516 70. Varshavsky, A. (2019) N-degron and C-degron pathways of protein degradation. Proc Natl
517 Acad Sci U S A 116 (2), 358-366.

518 71. Li, Z. et al. (2020) N-terminal acetylation stabilizes SIGMA FACTOR BINDING
519 PROTEIN 1 involved in salicylic acid-primed cell death. Plant Physiol 183 (1), 358-370.

520 72. Lai, Z. et al. (2011) Arabidopsis sigma factor binding proteins are activators of the
521 WRKY33 transcription factor in plant defense. Plant Cell 23 (10), 3824-41.

522 73. Lv, R. et al. (2019) Uncoupled expression of nuclear and plastid photosynthesis-associated
523 genes contributes to cell death in a lesion mimic mutant. Plant Cell 31 (1), 210-230.

524 74. Köhler, D. et al. (2015) Characterization of chloroplast protein import without Tic56, a
525 component of the 1-Megadalton translocon at the inner envelope membrane of chloroplasts.
526 Plant Physiol 167 (3), 972-990.

527 75. Waegemann, K. and Soll, J. (1996) Phosphorylation of the transit sequence of chloroplast
528 precursor proteins. J Biol Chem 271 (11), 6545-54.

529 76. Jarvis, J.A. et al. (1995) Solution structure of the acetylated and noncleavable mitochondrial
530 targeting signal of rat chaperonin 10. J Biol Chem 270 (3), 1323-31.

531 77. Lee, H.Y. et al. (2015) Arabidopsis serotonin N-acetyltransferase knockout mutant plants
532 exhibit decreased melatonin and salicylic acid levels resulting in susceptibility to an avirulent
533 pathogen. J Pineal Res 58 (3), 291-9.

534 78. Pietrzykowska, M. et al. (2014) The light-harvesting chlorophyll a/b binding proteins Lhcb1
535 and Lhcb2 play complementary roles during state transitions in Arabidopsis. Plant Cell 26 (9),
536 3646-60.

537 79. Serero, A. et al. (2001) Distinctive features of the two classes of eukaryotic peptide
538 deformylases. J Mol Biol 314 (4), 695-708.

539 80. Dirk, L.M. et al. (2001) Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast-
540 targeted enzymes in *Arabidopsis*. Plant Physiol 127 (1), 97-107.

541 81. Giglione, C. et al. (2003) Control of protein life-span by N-terminal methionine excision.
542 EMBO J 22 (1), 13-23.

543 82. Moon, S. et al. (2008) Rice peptide deformylase PDF1B is crucial for development of
544 chloroplasts. Plant Cell Physiol 49 (10), 1536-46.

545 83. Piatkov, K.I. et al. (2015) Formyl-methionine as a degradation signal at the N-termini of
546 bacterial proteins. Microb Cell 2 (10), 376-393.

547 84. Hoshiyasu, S. et al. (2013) Potential involvement of N-terminal acetylation in the
548 quantitative regulation of the ε subunit of chloroplast ATP synthase under drought stress. Biosci
549 Biotechnol Biochem 77 (5), 998-1007.

550 85. Schmidt, C. et al. (2017) Acetylation and phosphorylation control both local and global
551 stability of the chloroplast F(1) ATP synthase. Sci Rep 7, 44068.

552 86. Bienvenut, W.V. et al. (2011) Dynamics of post-translational modifications and protein
553 stability in the stroma of Chlamydomonas reinhardtii chloroplasts. Proteomics 11 (9), 1734-50.

554 87. Albanese, P. et al. (2020) How paired PSII–LHCII supercomplexes mediate the stacking of
555 plant thylakoid membranes unveiled by structural mass-spectrometry. Nat Commun 11 (1),
556 1361.

557 88. Deng, S. and Marmorstein, R. Protein N-Terminal Acetylation: Structural Basis,
558 Mechanism, Versatility, and Regulation. Trends in Biochemical Sciences.

559 89. Vener, A.V. et al. (2001) Mass spectrometric resolution of reversible protein
560 phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. J Biol Chem 276 (10),
561 6959-6966.

562 90. Gomez, S.M. et al. (2002) The chloroplast grana proteome defined by intact mass
563 measurements from liquid chromatography mass spectrometry. Mol Cell Proteomics 1 (1), 46-
564 59.

565 91. Turkina, M.V. et al. (2006) Environmentally modulated phosphoproteome of
566 photosynthetic membranes in the green alga Chlamydomonas reinhardtii. Mol Cell Proteomics
567 5 (8), 1412-25.

568 92. Baniulis, D. et al. (2009) Structure-function, stability, and chemical modification of the
569 cyanobacterial cytochrome b6f complex from Nostoc sp. PCC 7120. J Biol Chem 284 (15),
570 9861-9869.

571 93. Guskov, A. et al. (2009) Cyanobacterial photosystem II at 2.9-Å resolution and the role of
572 quinones, lipids, channels and chloride. Nat Struct Mol Biol 16 (3), 334-342.

573 94. Broser, M. et al. (2010) Crystal Structure of Monomeric Photosystem II from
574 Thermosynechococcus elongatus at 3.6-Å Resolution. J Biol Chem 285 (34), 26255-26262.

575 95. Yang, M.-k. et al. (2014) Proteogenomic analysis and global discovery of posttranslational
576 modifications in prokaryotes. Proc Natl Acad Sci U S A 111 (52), E5633-E5642.

577 96. Houtz, R.L. et al. (1992) Posttranslational Modifications in the Amino- Terminal Region of
578 the Large Subunit of Ribulose- 1,5-Bisphosphate Carboxylase/Oxygenase from Several Plant
579 Species. Plant Physiol 98 (3), 1170-1174.

580 97. Pfefferkorn, B. and Meyer, H.E. (1986) N-terminal amino acid sequence of the Rieske iron-
581 sulfur protein from the cytochrome b6/f-complex of spinach thylakoids. FEBS Lett 206 (2),
582 233-237.

583 98. Pierre, Y. et al. (1995) Purification and characterization of the cytochrome b_6 *f* complex
584 from *Chlamydomonas reinhardtii*. J Biol Chem 270 (49), 29342-29349.

585 99. Kouyianou, K. et al. (2012) Proteome profiling of the green sulfur bacterium
586 Chlorobaculum tepidum by N-terminal proteomics. Proteomics 12 (1), 63-7.

587 100. Misumi, O. et al. (2005) Cyanidioschyzon merolae genome. A tool for facilitating
588 comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiol 137 (2),
589 567-585.

590 101. Zhao, Y.-P. et al. (2019) Resequencing 545 ginkgo genomes across the world reveals the
591 evolutionary history of the living fossil. Nat Commun 10 (1), 4201.

592 102. Li, F.-W. et al. (2018) Fern genomes elucidate land plant evolution and cyanobacterial
593 symbioses. Nat Plants 4 (7), 460-472.

594 103. Zhang, L. et al. (2019) The water lily genome and the early evolution of flowering plants.
595 Nature 577 (7788), 79-84.

596 104. Grzela, R. et al. (2017) The C-terminal residue of phage Vp16 PDF, the smallest peptide
597 deformylase, acts as an offset element locking the active conformation. Sci Rep 7 (1), 11041.

598 105. Colaert, N. et al. (2009) Improved visualization of protein consensus sequences by
599 iceLogo. Nat Methods 6 (11), 786-7.

600

601

FIGURES

Key Figure 1. Main pathways leading to protein acetylation in plants. In photosynthetic organisms, NTA occurs in all cellular compartments where there is protein synthesis, except for mitochondria. In the cytoplasm, NTA is ensured by different NAT classes (A/B/C/E/F). NatE is made of NatA subunits (NAA10/NAA25) plus NAA50 (see Figure 4C in Ref.[88]). Although NAA40 has not yet been characterized in plants, *A. thaliana* histones H4 and H2A have similar N-termini as in fungi and humans suggesting similar function. Moreover, GNAT8 and GNAT9, initially identified as putative plastid NATs, were finally observed to localize to the cytoplasm and, in the case of GNAT8, the nucleus [25]. Seven additional GNATs (GNAT1/2/3/4/5/7/10) have been shown to localize to plastids and display dual NTA and KA activities [25]. GNAT6 is associated with both the outer side of the chloroplasts and the nucleus [25]. (1), (2), or (3) following a given catalyst indicates that the reaction is performed with various enzymes and the number refers to the order. SPP and TPP are plastid signal and transit peptide peptidase, respectively. Fo is N-formyl. This figure was created using BioRender (https://biorender.com/).

Figure 2. Summary of the physiological impact of different characterized plant NATs. Cartoon summarizing the main effects reported in the text. The same color code as in **Figure 1** is used to refer to the different NATs or NAAs.

Table 1. Compilation of the protein acetylation status in plastid major complexes in various photosynthetic organisms[a]

Function/ complex	Photosynthetic organism	NTAed subunits	UnNTAed subunits	% of the complex that contain NTAed proteins	Refs[b]
NDH	*A. thaliana*	Ndh45	NdhA/B/C/D/E/F/G/H/I/J/K/S/U	7%	[17, 23]
PSI	*A. thaliana*	PsaA	Psab/C/E/G/I/J/T	14%	[9, 17, 23]
	T. pseudonana	PsaD	PsaA/B/C/F/L	17%	[19]
PSII	*A. thaliana*	PsbA/B/C/D/F/H/J/L/Q/S	PsbE/I/K/M/N/O/T/Z	56%	[9, 17, 21-23, 25, 89, 90]
	P. sativum	PsbA/B/D/F/J/T	PsbE/H/L/M/T/Z	50%	[87][#]
	T. pseudonana	PsbC/D/F/L	PsbB/E	67%	[19]
	C. reinhardtii	PsbA/C/D	PsbE		[91]
	Nostoc	Psbf/J			[92]
	Synechococcus	PsbD/F/J			[93-95]
ATPase	*A. thaliana*	AtpA/B/E/H	AtpD/F/I	57%	[23]
	S. oleracea	AtpA/B/E			[85]
	T. pseudonana	AtpA	AtpD/H	33%	[19]
	C. reinhardtii	AtpA/B/E	AtpI	75%	[b]
RuBisCO	*A. thaliana*	Rbcl, RbcS		100%	[9, 23]
	M. polymorpha	RbcL			[96]
	T. pseudonana	RbcL			[19]
	C. reinhardtii	RbcL			[96]
LHC	*A. thaliana*	Lhc2.1/2.2/3.1/4/6, Lhcl3.1, Lhb1b2, Cp24	Lhc1/2/3, Lhb1b1, Cp26	62%	[17]
	P. sativum	Lhcb1/2/4.2/4.3/5	Lhcb3/6	71%	[87]
	C. reinhardtii	Lhcb4, Lhcbm4/6/9			[91]
	T. pseudonana	Lhcf2/5/7/9/11, Lhcr2, Lhcx2	Lhcr3/4/11/14, lhcf10, fcp2/10	50%	[19]
Ribosome	*A. thaliana*	Rpl2/5/15/34, Rps5/9/15	Rps2/3/4/7/8/11/12/13/14/15/16/17/18/19/21, Rpl1/2/3/4/6/12/13/14/16/19/20/21/22/23/32/33/36	17%	[17, 23]
	S. oleracea		rpl2/14/16/20/22/23/32/33/36;rps2/3/4/78/11/1/14/19	0%	[b]
	T. pseudonana	Rpl23	Rps2/3/4/7/8/11/12/13/14/15/16/17/18/19/21, Rpl1/2/3/4/6/12/13/14/16/19/20/21/22/23/32/33/36	17%	[19]
CYT b6f	*A. thaliana*		PetA/B/D/G/L/N	0%	[9, 17, 21, 23]
	S. oleracea		PetC		[97]
	T. pseudonana	PetB/C	PetA/D		[19]
	C. reinhardtii		PetA/C/M		[98]
	Nostoc	PetC			[92]
CLP	*A. thaliana*	ClpP1/3/5, ClpT1	ClpP6, Clpt2	67%	[17, 23]
	T. pseudonana	ClpB1			[19]
RNA polymerase	*A. thaliana*		RpoA/B/C1/C2	0%	[23]
	Zea maize		RpoB/C1/C2	0%	[b]
Other functions	*A. thaliana*	Tic214, DnaJ	MatK, Orf77		[23]
	C. tepidum	Q8KD44, Q8KFS3, Q8KE86, Q8KAU3, Q8KAL6			[99]
	T. pseudonana		DnaK, GroL, ThiG,Ycf39, Ycf46	0%	[19]

[a] Proteins can arise from both plastid and nuclear genomes. References are indicated. A database compiling *A. thaliana* and other land plant data is available at NTerDB at

https://energiome.i2bc.paris-saclay.fr/ [18]. NDH is for NADPH dehydrogenase, CYT for Cytochrome and CLP for ATP-dependent caseinolytic protease.

[b] See table available at https://www.i2bc.paris-saclay.fr/spip.php?article1261. The corresponding references are detailed at https://www.i2bc.paris-saclay.fr/spip.php?article1262.

Table 2. Plant NAA and NAT complexes[a]

Cell localization	Organism			*S. cerevisiae* (Yeast)	*H. sapiens* (Human)	*A. thaliana*	*O. sativa* (Rice)
	Complex	Subunit		Accession number	Accession number	Gene No	Gene No
Cytosol	NatA/E	**NAA10**		P07347	P41227	At5g13780	Os04g0635800
		NAA11[b]		*NR*	Q9BSU3	*NR*	*NR*
		NAA15[c]		P12945	Q9BXJ9	At1g80410	Os01g0617500
		NAA16[c,d]		*NR*	Q6N069	*NR*	*NR*
		HYPK[c]		*NR*	Q9NX55	At3g06610[f]	Os03g61680[f]
		NAA50[e]		Q08689	Q9GZZ1	At5g11340	P0410E03.23
	NatB	**NAA20**		Q06504	P61599	At1g03150	Os11g32280
		NAA25[c]		Q12387	Q14CX7	At5g58450	Os05g0345400
	NatC	**NAA30**		Q03503	Q147X3	At2g38130	Os11g32280
		NAA35[c]		Q02197	Q5VZE5	At2g11000	P0689B12.25
		NAA38[c]		P23059	O95777	At4g18372[f]	Os06g0677600[f]
	NatD	**NAA40**		Q04751	Q86UY6	At1g18335	Os05g0387800
	NatF	**NAA60**		*NR*	Q9H7X0	At5g16800	B1053A04.3
	NatH	**NAA80**		*NR*	Q93015	*NR*	*NR*
	-	**GNAT8**		*NR*	*NR*	At2g39020	Os03g05710
	-	**GNAT9**		*NR*	*NR*	At2g04845	OsJ_09757
Plastid	NatG[g]	**NAA90**	**GNAT1**	*NR*	*NR*	At1g26220	Os08g0102000
			GNAT2/NSI	*NR*	*NR*	At1g32070	Os05g0481000
			GNAT3	*NR*	*NR*	At4g19985	LOC4339320
		NAA70	**GNAT4**	*NR*	*NR*	At2g39000	Os02g0806000
			GNAT5	*NR*	*NR*	At1g24040	Os04g0465500
			GNAT6	*NR*	*NR*	At2g06025	LOC4331807
			GNAT7	*NR*	*NR*	At4g28030	B1015H11.141
			GNAT10	*NR*	*NR*	At1g72030	Os07g0264800

[a] Each entry is indicated for all NAAs from *S. cerevisiae*, *H. sapiens*, *A. thaliana*, and *O. sativa*.
[b] Duplicated version, alternative to NAA10
[c] Auxiliary subunit, without catalytic activity
[d] Duplicated version, alternative to NAA15
[e] Inactive in *S. cerevisiae*, where it acts as sheer auxiliary subunit
[f] Remains to be validated experimentally
[g] No evidence for the formation of a complex built with various GNAT subunits. Globally refers to plastid NTA (NAA70 and NAA90 families, see **Box 1**).
NR, not relevant, means that the corresponding orthologous component was not identified.

BOXES

Box 1: Evolutionary boundaries of NAAs across the plant kingdom

The Nat machinery varies across the various clades and orders of the plant kingdom, which includes algae and land plants (**Figure IA**). NAAs all share the dedicated GNAT family structural domain involved in acetyl-CoA binding (domain architecture ID 11418877) and they are usually associated with the RimI-conserved domain (COG0456). **Figure IA** is a phylogram constructed from 179 NAA orthologs found in 13 representative species for which robust genome and/or proteome annotation were available in each clade. The updated phylogram adds new key species such as the red alga *Cyanidioschyzon merolae* [100], the living fossil gymnosperm maidenhair tree *Ginkgo biloba* [101], the small fern fairy moss *Azolla filiculoides* [102], and the magnoliids water lily *Nymphaea colorata* [103] (**Figure IB**). Remarkably, all plants, from algae to dicots, show very similar features. First, they all display NAA10/20/30/40. NAA50 could not be retrieved in red algae, whereas NAA60 could not be identified in algae. Second, GNAT4 orthologs could be retrieved in all species (**Figure IB**). As suggested in [25], it is clear that GNAT5/6/7 and GNAT10 are closely related to GNAT4 (NAA70s) and might have resulted from progressive gene duplication, as observed in many plant genomes. This does not necessarily imply that they do not display their own specificity or exhibit differences, as illustrated in **Figure ID**. Nevertheless, not all representatives of the kingdom contain all members. Red algae, which have a reduced proteome, contain only GNAT4; all GNAT4-related proteins were included in the same subfamily (NAA70; **Table 2**). Similarly, as GNAT1/2/3 are closely related, they were proposed to belong to the same subfamily, NAA90 (**Table 2**). At least one NAA90 family ortholog is present in all representative members of the plant kingdom. GNAT8, previously predicted to be a plastid NAT but finally described as a cytosolic protein [25], has only been identified in land plants, in contrast to GNAT9, which has been identified

in all plants except red algae. GNAT9 was reported to more closely resemble the RimJL family (COG1670) [25].

Therefore, in plants, the minimal machinery consists of one copy of cytosolic NAA10/20/30/40 and of the plastid-targeted subfamilies NAA70 and NAA90. This is the case for red algae, which lack NAA50, NAA60 as well as both GNAT8 and GNAT9.

Box 1, Figure I. Photosynthetic organisms express a large array of NATs, including plastid-specific isoforms. Most data are available at UniProt (https://www.uniprot.org/proteomes/), the algal genomics resource at the JGI (https://mycocosm.jgi.doe.gov/phycocosm/home) and, when missing, dedicated resources (e.g., http://cyanophora.rutgers.edu/cyanophora).

(A) Organisms used to sample the green NAA landscape and their positions with respect to eukaryotic evolution. 13 organisms were selected to cover and evenly sample the diversity of photosynthetic eukaryotic proteomes. The figure with the tree of plants according to the Haeckel tree of life (1866). Each proteome was screened from data mostly available in UniProt reference proteomes (https://www.uniprot.org/proteomes/). PSI-BLAST analysis was systematically performed at NCBI (https://blast.ncbi.nlm.nih.gov/) to identify annotated homologs. TBLASTN was also searched as required. (B) 176 sequences were selected to represent the sequence diversity of the catalytic subunits of NATs among photosynthetic species. The sequences were aligned, and the bootstrap tree was constructed as previously described [104]. Internal values labeled on each node record the stability of the branch over the 1,000 bootstrap replicates. The figure displays a simplified representation of the tree. All data, distance, and entry numbers are available upon request. The NAA-60 nomenclature is similar to that [1] (see also **Table 2**).

Box 2. Assessing NAT substrate specificity

The development of the global acetylation profiling (GAP) assay has been instrumental for assessing NTA activity and substrate specificity of plant NATs [55]. The GAP assay is based on the expression of putative NAAs in *E. coli*, taking advantage of the low level of NTAed proteins in *E. coli* and exploiting a dedicated protocol for the enrichment, identification, and quantification of NTAed N-termini. Using this assay, expression of AtNAA10 in *E. coli* [55] recapitulated plant selectivity as deduced directly by comparison of wild-type acetylation and NatA knockdown plant lines [13]. **Figure I** illustrates how both approaches - which compile data from dozens of acetylated and non-acetylated proteins - are complementary and reveal that: (i) Ser-starting proteins are the best substrates for AtNAA10 (GAP assay, **Figure IB**); (ii) Ala-starting proteins are the most sensitive to AtNatA downregulation (**Figure IA**); and (iii) there is high counter-selection against Met-starting proteins. Why *A. thaliana* AtNAA15 does not contribute to AtNAA10 selectivity as much as in other organisms is currently unclear.

N-terminomics global analysis based on *A. thaliana* lines exhibiting reduced NatB levels [14] revealed a clear preference for AtNatB towards N-termini retaining their iMet followed by Asp and Glu and, to a smaller extent, Asn (**Figure IC**).

Additionally, a GAP assay with AtNAA50 revealed it had broad substrate specificity for N-termini, partially overlapping with human NAA50 substrate specificity (see online supplemental information **Figure S1A**) [39]. GAP assaying of AtNAA60 [52] revealed substrate specificity comparable to that of HsNAA60, with a preference for an iMet and a highly variable position 2 (see online supplemental information **Figure S1B**).

GNATs have relaxed specificity. This relaxed NTA substrate specificity was mainly dependent on the amino acids at positions 1, 2, and 3 (**Figure ID**). For instance, all of the six most active plastid AtGNATs were very efficient with N-termini starting with an iMet but showed clear differences induced by the position 2 amino acid. GNAT1 was the only GNAT

that efficiently acetylated Met-Met substrates, and GNAT10 was more prone to acetylating N-termini starting with a Met, showing similar substrate specificity to NatC/E/F and to a lesser extent NatB-like substrates. GNAT2, 4, 5, and 7 acted more efficiently on NatA-like substrates, albeit with different impact depending on the amino acid at position 2 (**Figure I D**).

Box 2, Figure I. Hints towards the substrate specificity of the main *A. thaliana* NATs. Protein logo representations in panels A-C were displayed using iceLogo [105]. Positive and negative sets were used to compare NTAed *vs.* non-NTAed positions in the mutant plant lines or NTAed *vs.* non-NTAed positions recovered from the GAP assay. Unlike a classic logo, iceLogo reveals favored and disfavored positions whilst masking positions that are neutral as equally occurring in both sets. (A) Specificity of AtNatA as deduced from (NAA10+NAA15) knockdown in Arabidopsis. N-terminomics data from *amiNAA10* and *amiNAA15* knockdown plant lines (available in [13]). The positive set includes 108 N-termini, which were significantly decreased upon reduced expression of NatA (i.e., >20%) in either or both lines; the negative set includes 220 N-termini modified below 5%. (B) AtNAA10 specificity as deduced from the GAP assay. Data are from [55]. The same cutoffs as in (A) were used to define the positive (45 N-termini) and negative sets (92). (C) Same as (A) but with NatB in *naa20-1* and *naa20-2* T-DNA insertion mutant lines exhibiting reduced NatB expression [14]. 28 positives *vs.* 57 negatives were used to build the iceLogo. Data are from [14]. (D) Unique and redundant substrate specificity of plastid AtGNATs. The data were compiled from [25]. The data of the acetylation yield resulting from global acetylome profiling of the six more significant plastid AtGNATs are reported in each column. All protein substrates starting with the same first two amino acids are considered as identical, and the upper value of acetylation yield is reported. Amino acids at position two are grouped according to the chemical properties of their side-chains. Acetylation yield is reported and associated with a color code where red and blues indicate the highest and lowest levels, respectively. The first two amino acids of NAT substrates

740 are reported in the first two columns. A value of x means x% of the retrieved *E. coli* proteins
741 starting with amino acids y are NTAed if GNATz is expressed.
742

GLOSSARY

Acetylation yield: one single protein may feature *in vivo* both unmodified (free) and N-acetylated termini resulting in partial *in vivo* NTA. The acetylation yield corresponds to the NTAed/(Free+NTAed) fraction. A fully NTAed protein displays a 100% yield whereas a fully unmodified protein has a 0% yield. The yield value ranges 0 to 100.

General control non-repressible 5 (GCN5)-related N-acetyltransferases (GNATs): a superfamily of N-acetyltransferases using acetyl CoA as the acetyl donor and a number of small or larger molecules as acceptors.

Global acetylation profiling (GAP) assay: N-terminomics analysis of *E. coli* proteins after expression of a catalytically active NAA. Allows unbiased quantification and assessment of the NTA yields of dozens of random substrates simultaneously. Reveals substrate selectivity of the studied NAA by comparing NTAed to non-NTAed proteins.

Huntingtin interacting protein K (HYPK): a protein that protects against Huntingtin polyQ-mediated apoptosis and regulates the NatA/E complex.

N-Acetylated at the aminoterminus (NTAed): refers to a protein that is acetylated on its N-terminal amino group. This NTA can be partial (see Acetylation yield)

N-terminal α-acetylation (**NTA**): co- and post-translational modification consisting of the addition of an acetyl group to the positively charged free α-amino group of the first amino acid of proteins.

N-terminal methionine excision process (NME): co-translational modification involving deformylation by peptide deformylases (PDFs) in mitochondria and chloroplasts and methionine aminopeptidases (MetAPs) in all cellular compartments where protein synthesis occurs. NME consists of the sequential trimming of the first Met (iMet or FormylMet) from nascent chains.

N-terminal acetyltransferase (NAT): enzyme belonging to the GNAT superfamily that catalyzes NTA, composed of a catalytic subunit (NAA) that can be associated with one or more auxiliary subunits.

Light-harvesting complex (LHC) (or antenna supercomplex): a complex formed from a large number of protein subunits and chlorophyll molecules embedded in the thylakoid membrane of plants. It collects sunlight and transfers it to photosystems.

Lysine ε-acetylation (KA): reversible post-translational modification involving the transfer of an acetyl group from a donor (e.g., acetyl CoA or acetyl phosphate) to ε-amino groups of lysines.

Lysine ε-acetyltransferase (KAT): enzyme controlling KA. KATs belong to four families: P300/CBP, MYST, GNAT, or YoJ.

Photosystem II (PSII): multi-subunit pigment-protein complex embedded in the thylakoid membranes of all oxygenic photosynthetic organisms that catalyzes the oxidation of water using solar energy. Some microorganisms like Chlorobi (*Chlorobium tepidum*) do not use PSII.

Plastid: organelle found in the cells of plants, algae, and some other eukaryotic organisms (e.g., parasites such as Apicomplexa) originating from cyanobacteria endosymbiosis into other primitive eukaryotic cells.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO): a highly abundant stromal complex made of eight large catalytic (RbcL) subunits and eight small regulatory (RbcS) subunits. It is primarily involved in CO_2 assimilation into most organic compounds. It also displays photorespiration activity.

Serotonin acetyltransferase (SNAT): an enzyme that catalyzes the rate-limiting step of melatonin synthesis, the N-acetylation of serotonin to N-acetylserotonin.

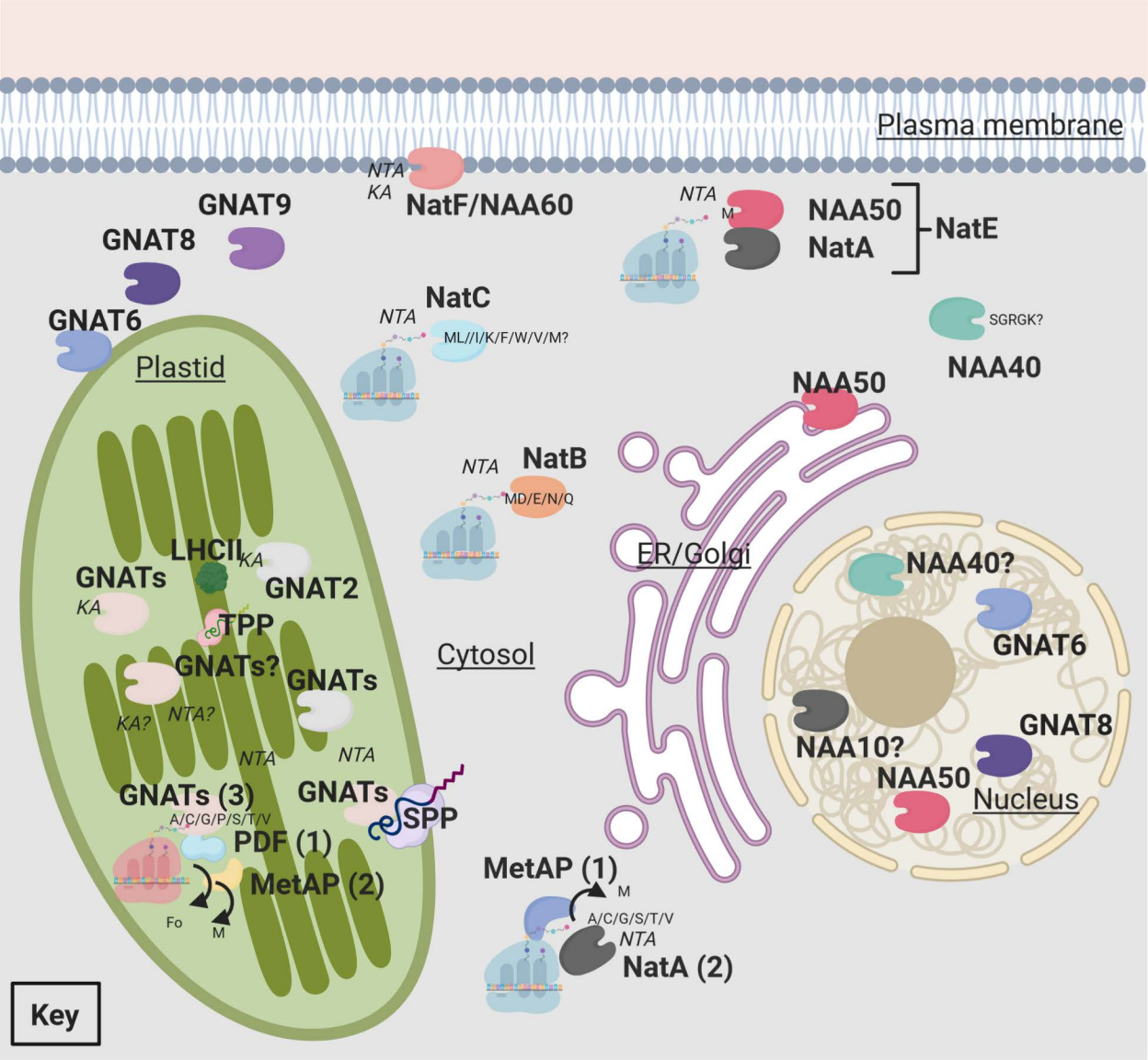
HIGHLIGHTS

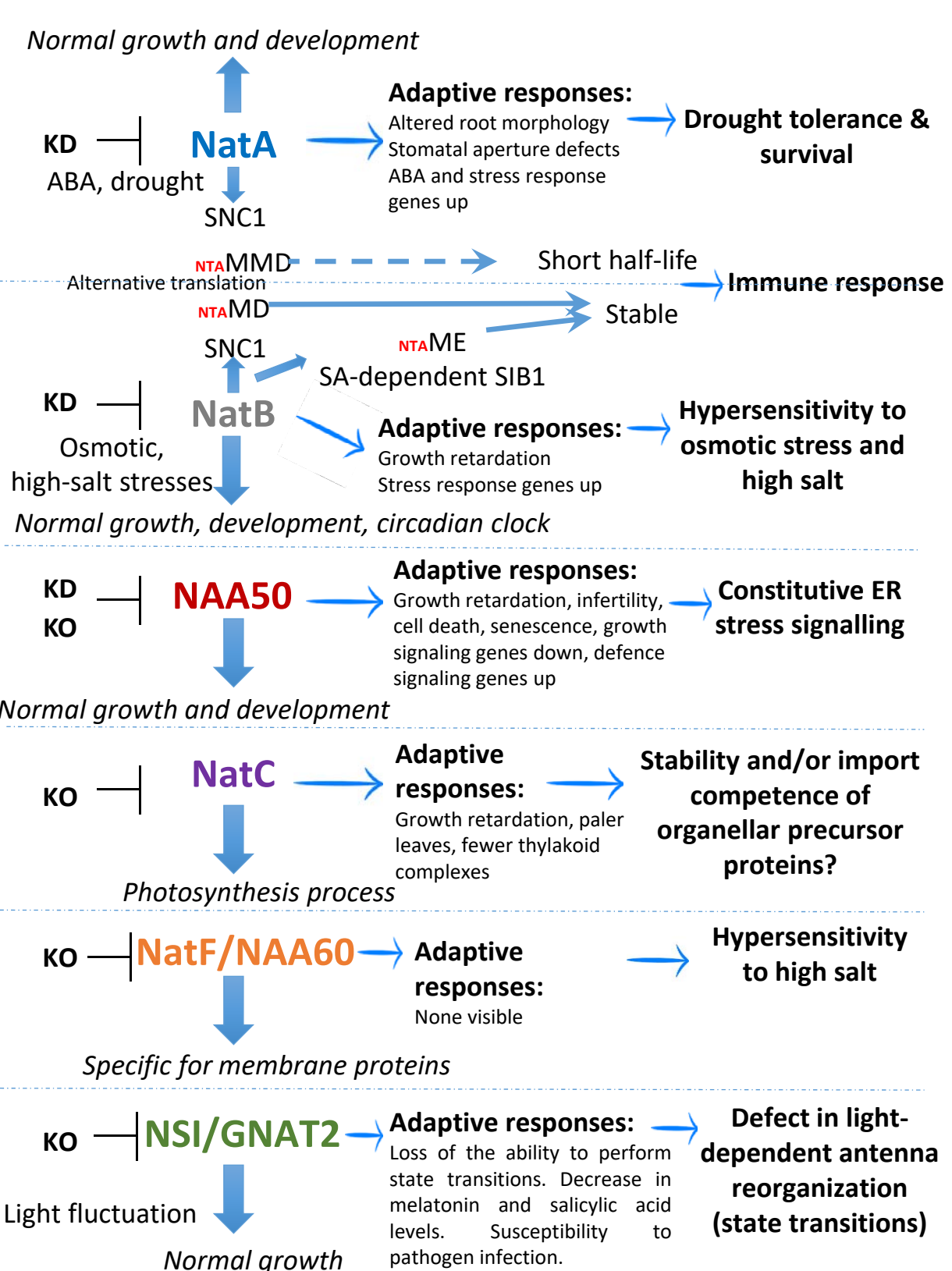
- The identification and characterization of the cytosolic and plastid NAT machinery has revealed the importance of NTA in plant development, responses to abiotic stresses, immunity, and protein translocation and stability.
- A unique plastid NAT machinery has now been identified and characterized.
- Several NATs show dual KA and NTA activities, including plastid forms.
- There is strong conservation of the entire machinery across the Viridaeplanta phylum.
- Plastid photoxygenic autotrophy-dedicated machineries such as RuBisCO, LHC, or PSII display significantly more frequent NTA of their subunits.

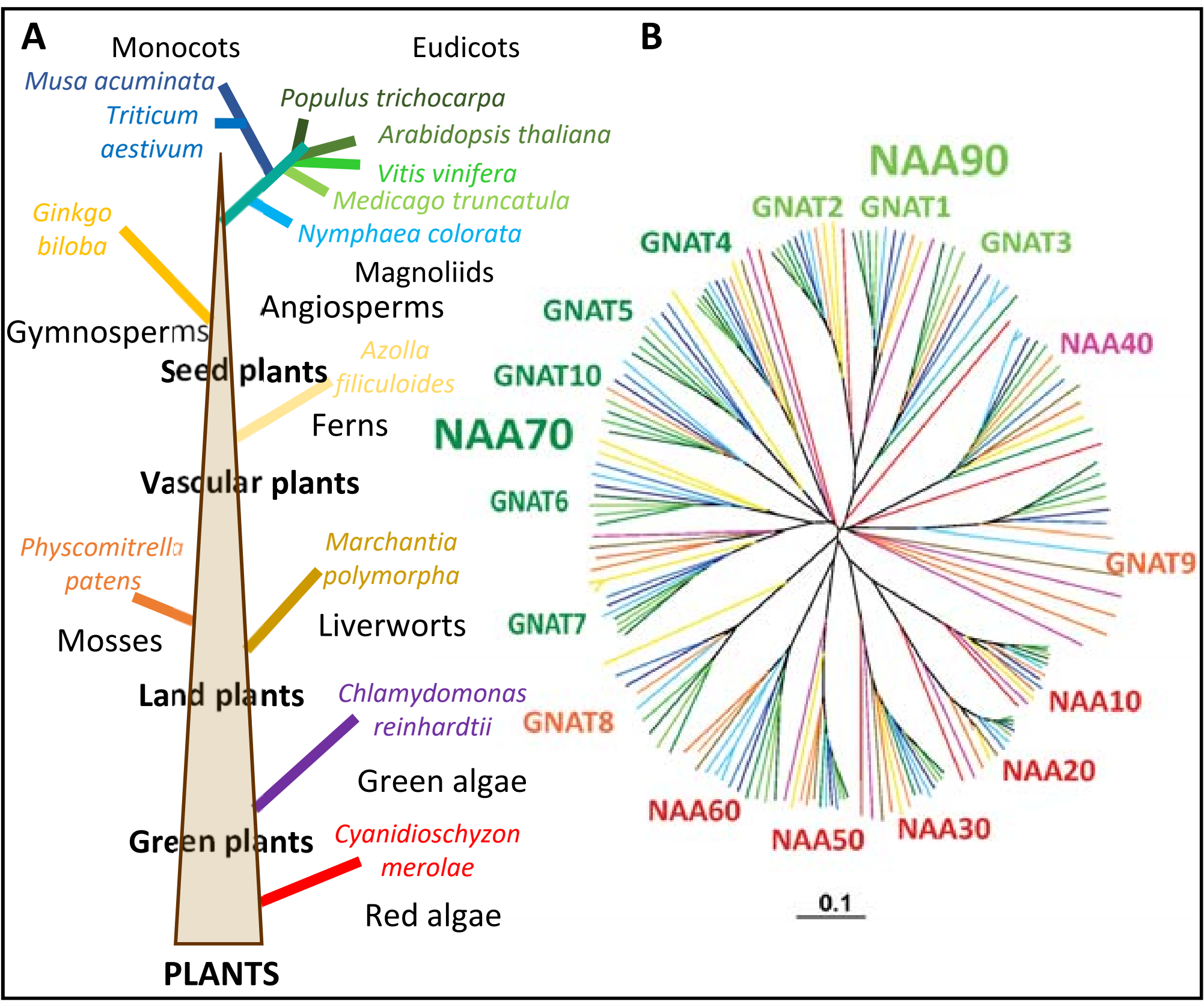
OUTSTANDING QUESTIONS

- Why does *A. thaliana* AtNAA15 not contribute to AtNAA10 selectivity as much as in other organisms?
- What is the real nature of the NatA/E complex in photosynthetic organisms? Do NatA and NatE independently exist or do they actually belong to the same complex?
- Are the free Naa10 or Naa50 catalytic subunits of photosynthetic organisms active *in vivo* like in metazoans? Do they exhibit KA, NTA, or both activities, and in which compartment(s)? Characterization of the putative plant HYPK will help to clarify the dialogue between NAA50 and other NatA subunits.
- What molecular mechanisms underpin the effect of cytosolic NATs on abiotic and biotic stress responses? Are NTA dynamics influenced by other factors such as stresses or dialogues with other modifications?
- How do the different C-termini of AtNAA60 influence membrane localization?
- Does NatA- and NatB-induced NTA trigger generally opposing effects on protein turnover?
- Do plant NAA40s also specifically acetylate - like in fungi and humans - histone variants H2A and H4 starting with SGRK? Are plant NAA40s also both nuclear and cytosolic?
- Which proteolytic machineries are involved in NatA-dependent protein turnover, including SNC1?
- What is the impact of cytosolic and plastid NATs on N-degrons?
- How and why do NAAs achieve both KA and NTA activities?
- What are the specificity determinants for KA activity?
- What does NTA add to the biology of the major photooxygenic autotrophy complexes in plants including LHC, PSII, and RuBisCO?
- Why are so many plastid proteins only partially acetylated?

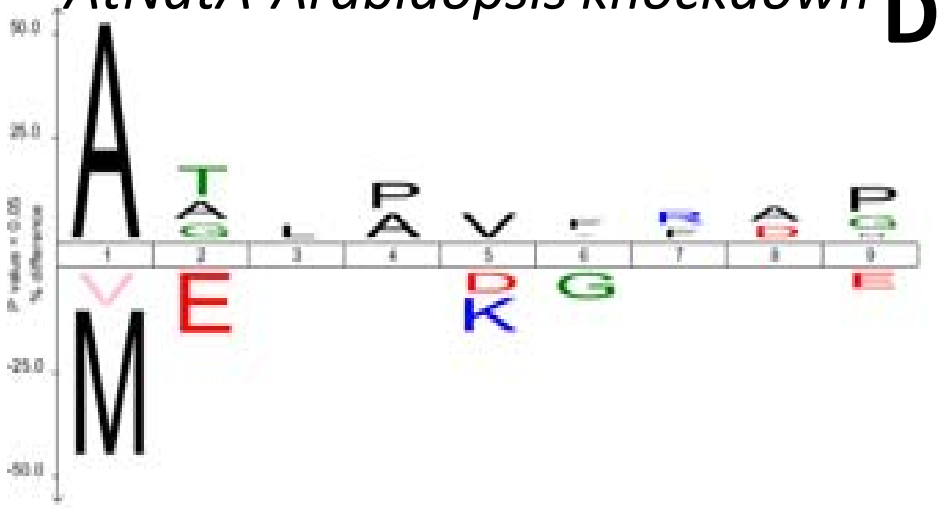
- Why are there so many different GNATs in plastids? Is this due to redundancy, the specific functions of each, or both?
- Do plastid GNATs act within higher complexes together or with auxiliary proteins like the major cytosolic NATs? Which ones act co-translationally on plastid-encoded proteins and post-translationally on imported proteins?


Plasma membrane
NTA
KA
NatF/NAA60
NTA
M
NAA50
NatA
NatE
GNAT9
GNAT8
GNAT6
NatC
NTA
ML//I/K/F/W/V/M?
SGRGK?
NAA40
Plastid
NAA50
NTA
NatB
MD/E/N/Q
ER/Golgi
LHCII
KA
GNATs
KA
GNAT2
TPP
GNATs?
GNATs
KA?
NTA?
Cytosol
NAA40?
GNAT6
GNAT8
NAA10?
NAA50
Nucleus
NTA
NTA
GNATs (3)
A/C/G/P/S/T/V
GNATs
SPP
PDF (1)
MetAP (2)
Fo
M
MetAP (1)
M
A/C/G/S/T/V
NTA
NatA (2)
Key

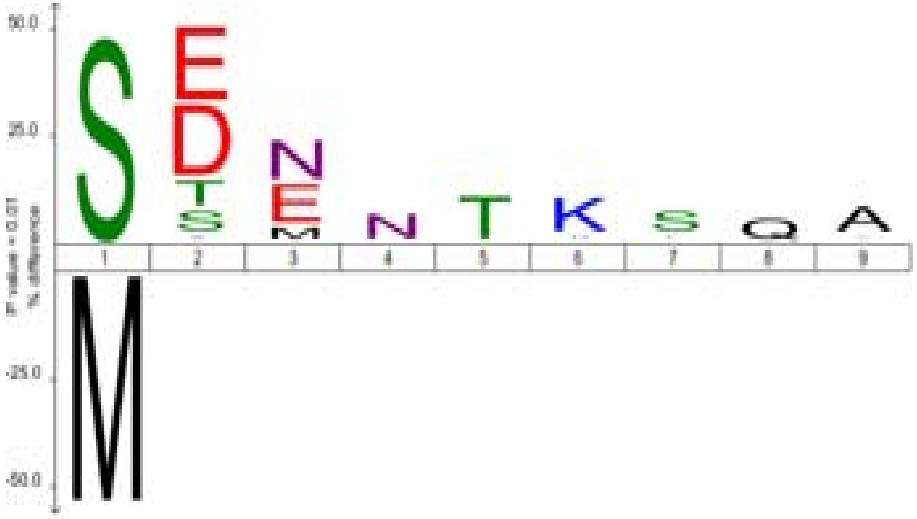

Normal growth and development
KD
ABA, drought
NatA
SNC1
Adaptive responses:
Altered root morphology
Stomatal aperture defects
ABA and stress response genes up
Drought tolerance & survival
NTAMMD
Alternative translation
Short half-life
Immune response
NTAMD
Stable
NTAME
SNC1
SA-dependent SIB1
KD
Osmotic, high-salt stresses
NatB
Adaptive responses:
Growth retardation
Stress response genes up
Hypersensitivity to osmotic stress and high salt
Normal growth, development, circadian clock
KD
KO
NAA50
Adaptive responses:
Growth retardation, infertility, cell death, senescence, growth signaling genes down, defence signaling genes up
Constitutive ER stress signalling
Normal growth and development
KO
NatC
Adaptive responses:
Growth retardation, paler leaves, fewer thylakoid complexes
Stability and/or import competence of organellar precursor proteins?
Photosynthesis process
KO
NatF/NAA60
Adaptive responses:
None visible
Hypersensitivity to high salt
Specific for membrane proteins
KO
NSI/GNAT2
Adaptive responses:
Loss of the ability to perform state transitions. Decrease in melatonin and salicylic acid levels. Susceptibility to pathogen infection.
Defect in light-dependent antenna reorganization (state transitions)
Light fluctuation
Normal growth


A
Monocots
Eudicots
Musa acuminata
Triticum aestivum
Populus trichocarpa
Arabidopsis thaliana
Vitis vinifera
Medicago truncatula
Nymphaea colorata
Ginkgo biloba
Magnoliids
Angiosperms
Gymnosperms
Seed plants
Azolla filiculoides
Ferns
Vascular plants
Physcomitrella patens
Marchantia polymorpha
Liverworts
Mosses
Land plants
Chlamydomonas reinhardtii
Green algae
Green plants
Cyanidioschyzon merolae
Red algae
PLANTS
B
NAA90
GNAT2
GNAT1
GNAT4
GNAT3
GNAT5
NAA40
GNAT10
NAA70
GNAT6
GNAT9
GNAT7
NAA10
GNAT8
NAA20
NAA60
NAA50
NAA30
0.1

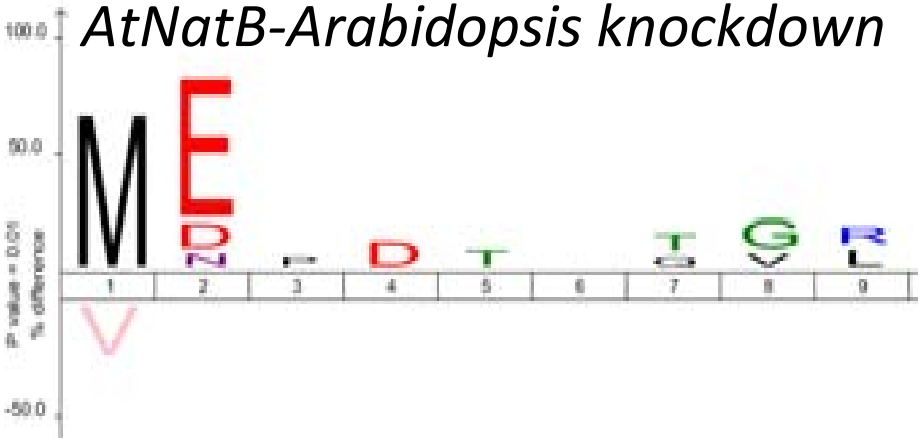
A *AtNatA-Arabidopsis knockdown*



B *AtNAA10-GAP*



C *AtNatB-Arabidopsis knockdown*



D

GNAT		2	4	5	6	7	10
aa1	aa2						
Ala	AGST	100	44	30	25	35	30
	DE	31	99	95	65	50	29
	K	34	57	76	52	26	
	ILMV	97	100	99	100	37	97
	NQ	98	100	25		5	4
	FY		99	77	14	27	
Gly	T		19				
Ile	K						6
Met	GST		12	95	45	76	100
	DE	61	100	81	99	71	33
	HK	55	27	35	94	96	55
	ILVM	92	91	97	100	94	94
	NQ	100	100	50	100	99	50
	FY	27	18	100	100	56	93
Pro	L					90	
	S					5	
Ser	ACST	46	100	89	88	67	52
	DE	94	59	15	20	65	27
	K	6	29		9	16	17
	ILMV	29	100	62	65	11	14
	NQ	83	99	99	55	30	77
	F	7		34	31		
Thr	AST	+	+		61		
	DE	32	23	98	20	82	7
	KR		12			86	
	ILMV	97	96	98	100	58	16
	NQ	100	99	5	24	99	
	F			+	34	15	28
Val	AS	13	+	+	99	96	
	D	94	95		87	87	96
	K					78	
	WY	48	12	12	5	58	9

Supplemental Information

Evolution-driven versatility of N-terminal acetylation in photoautotrophs

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

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

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

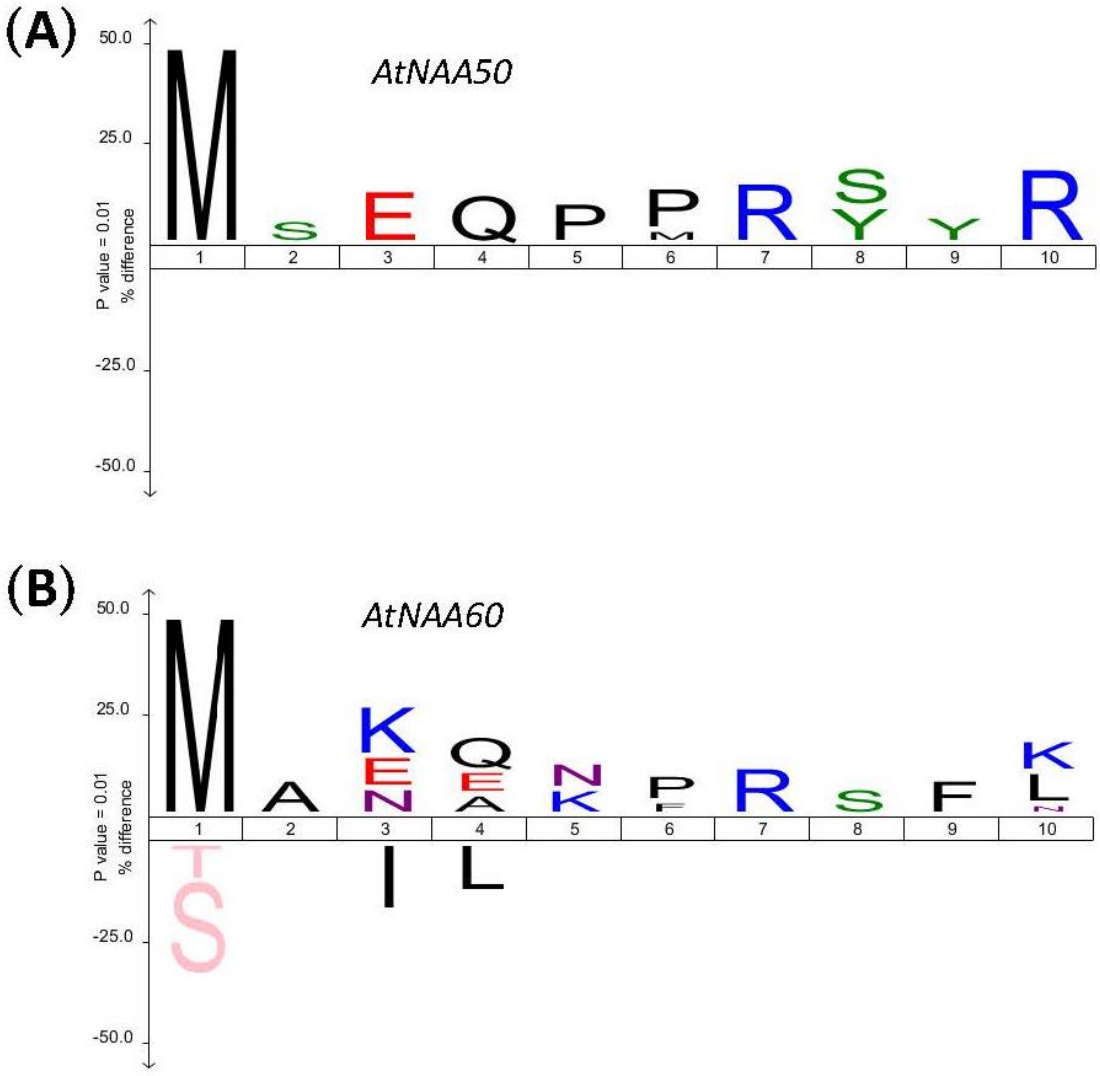


Figure S1. Substrate selectivity of *Arabidopsis* NAA50 and NAA60. GAP assays were performed with **(A)** AtNAA50 and **(B)** NAA60. Data are from [39, 52] The same cutoffs as in **Figure 2** were used to define the positive (19 and 68 N-termini, respectively) and negative (106 and 130, respectively) sets in either panel.