

RESEARCH ARTICLE

Molecular identification and functional characterization of the first N α -acetyltransferase in plastids by global acetylome profiling

Trinh V. Dinh^{1*}, Willy V. Bienvenut^{2*}, Eric Linster^{1,3}, Anna Feldman-Salit⁴, Vincent A. Jung², Thierry Meinne², Rüdiger Hell¹, Carmela Giglione² and Markus Wirtz¹

¹ Department of Plant Molecular Biology, Centre for Organismal Studies, University of Heidelberg, Heidelberg, Germany

² Institute of Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Gif-sur-Yvette, France

³ Hartmut Hoffmann-Berling International Graduate School, University of Heidelberg, Heidelberg, Germany

⁴ Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies gGmbH, Heidelberg, Germany

Protein N α -terminal acetylation represents one of the most abundant protein modifications of higher eukaryotes. In humans, six N α -acetyltransferases (Nats) are responsible for the acetylation of approximately 80% of the cytosolic proteins. N-terminal protein acetylation has not been evidenced in organelles of metazoans, but in higher plants is a widespread modification not only in the cytosol but also in the chloroplast. In this study, we identify and characterize the first organellar-localized Nat in eukaryotes. A primary sequence-based search in *Arabidopsis thaliana* revealed seven putatively plastid-localized Nats of which AT2G39000 (AtNAA70) showed the highest conservation of the acetyl-CoA binding pocket. The chloroplastic localization of AtNAA70 was demonstrated by transient expression of AtNAA70:YFP in *Arabidopsis* mesophyll protoplasts. Homology modeling uncovered a significant conservation of tertiary structural elements between human HsNAA50 and AtNAA70. The *in vivo* acetylation activity of AtNAA70 was demonstrated on a number of distinct protein N α -termini with a newly established global acetylome profiling test after expression of AtNAA70 in *E. coli*. AtNAA70 predominately acetylates proteins starting with M, A, S and T, providing an explanation for most protein N-termini acetylation events found in chloroplasts. Like HsNAA50, AtNAA70 displays N α -acetyltransferase activity on three internal lysine residues. All MS data have been deposited in the ProteomeXchange with identifier PXD001947 (<http://proteomecentral.proteomexchange.org/dataset/PXD001947>).

Received: January 15, 2015

Revised: March 20, 2015

Accepted: April 30, 2015

Keywords:

Arabidopsis thaliana / AtNAA70 / Chloroplast / N α -acetyltransferase / Plant proteomics



Additional supporting information may be found in the online version of this article at the publisher's web-site

Correspondence: Dr. Markus Wirtz, Department of Plant Molecular Biology, Centre for Organismal Studies, University of Heidelberg, Im Neuenheimer Feld 360, 69120 Heidelberg, Germany
E-mail: markus.wirtz@cos.uni-heidelberg.de
Fax: +49 6221 54-5859

Abbreviations: Acetyl-CoA, acetyl coenzyme A; AtNAA70, AT2G39000; EYFP, enhanced yellow fluorescent protein; GAP, global acetylome profiling; His₆-MBP, 6xhistidine-tag-maltose binding protein; iMet, Initial methionine; Kat, N α -

1 Introduction

N α -terminal acetylation (NTA) is one of the most common protein modifications in eukaryotes, occurring on 80–90% of

acetyltransferase; **Nat**, N α -acetyltransferase; **NTA**, N-terminal acetylation; **SCX-LC**, strong cation exchange-LC; **SILProNAQ**, stable isotope labeling protein N-terminal acetylation quantitation

*These authors contributed equally to this work.

Colour Online: See the article online to view Figs. 1–5 in colour.

cytosolic human proteins and 50–70% of cytosolic yeast proteins [1, 2]. However, the biological role of this modification has remained enigmatic. In few cases, NTA has been shown to affect protein–protein interaction and subcellular targeting [1, 3, 4]. NTA has been suggested to stabilize protein by blocking ubiquitination mediated degradation in the N-end rule pathway [5]. On the contrary, some acetylated N-termini can be recognized by Doa10, a ubiquitin ligase that targets proteins for proteasome degradation in yeast [6]. Very recently, also aggregation has been linked to N-terminal acetylation status of proteins [7].

NTA is catalyzed by N^α-acetyltransferases (Nats) which transfer the acetyl group from acetyl-coenzyme A (acetyl-CoA) to the α-amino group of the N-terminal amino acid of a protein and is considered to be irreversible. The acetylated amino acid may be the initiator methionine (iMet) or the first residue after the removal of iMet by methionine aminopeptidases [8]. This co-translational modification occurs when 10–70 residues have left the exit tunnel [9, 10]. Eukaryotic Nats usually consist of a catalytic subunit and sometimes, an auxiliary subunit suggesting to favor interaction with other Nat binding partners and/or to anchor the complex to the ribosome [1, 11, 12], e.g. NatA consists of the catalytically active NAA10 and the auxiliary subunit NAA15. At present five catalytic subunits of Nats (ScNAA10–ScNAA50) are known from yeast and six from humans (HsNAA10–HsNAA60) of which NAA10 to NAA50 orthologous are conserved with respect to amino acid sequence, cytosolic localization and substrate specificity [1, 13, 14].

In plants, besides the co-translational acetylation of cytosolic proteins, NTA is reported to be a widespread modification of chloroplastic proteins from both origins, i.e. those deriving from the plastid genome expression and plastoribosomes translation (e.g. RbcL and CP43; see <http://www.isv.cnrs-gif.fr/recherche/tm/maturation/images/chloro.html> and [15, 16]) and the nuclear-encoded proteins which are further imported in the plastid [17–20]. This implies that NTA occurs not only co-translationally for plastid-encoded proteins as in the cytosol but also post-translationally for plastid-imported proteins of photosynthetic organisms [20]. Furthermore, the post-translational acting Nat(s) might also (in part) be responsible for the N-terminal acetylation of plastid-encoded proteins. Interestingly, NTA of the chloroplastic ε subunit of ATP synthase protein was evidenced to regulate its stability during drought stress [21]. Similarly, the stability of the plastid proteome is dependent on the NTA status and other N-terminal modifications in the green algae *Chlamydomonas reinhardtii* [22]. Despite the importance of NTA in the chloroplast for regulation of protein turnover, the NTA machinery of chloroplastic proteins in general is uncharacterized (i.e. co- and/or post-translational NTA). Only cytosolic AtNAA20 and AtNAA30 from *Arabidopsis thaliana* have so far been functionally analyzed in higher plants [23, 24]. In this study, a homology based in silico search of the Arabidopsis genome revealed 25 putative Nats, of which seven are predicted to be localized in the chloroplast. Subcellular localization

in Arabidopsis protoplasts and the use of a new global acetylome profiling (GAP) approach clearly demonstrate that AtNAA70 displays indeed N^α-acetyltransferase activity. The substrate specificity of AtNAA70 was even broader than substrate specificities of cytosolic metazoan Nats. Pairwise alignment and modeling of AtNAA70 to cytosolic human NAA50 (HsNAA50) allows identification of the catalytically active site of AtNAA70 and reveals a remarkable conservation of the tertiary structure between post-translationally acting AtNAA70 and co-translationally cytosolic metazoan Nats. Interestingly, AtNAA70 possesses, like HsNAA50 and HsNAA10, auto-Lys^ε acetylation (Kat) activity.

2 Materials and methods

2.1 Three-dimensional structure modeling of N^α-acetyltransferases from Arabidopsis

Local sequence pairwise alignments of putative Arabidopsis Nat encoded by loci AT2G39000 (Q9ZV08) against human HsNAA50 (2PSW, Q9GZZ1) were performed with BLASTP 2.2.30 [25]. A tertiary structure was defined for the regions at residues 63–112 and 195–254 of AT2G39000 by comparative modeling against the resolved quaternary structure of 2PSW with MODELLER9.1 (applied settings: “automodel.very_fast”, [26]). Conservation analysis and visualization was performed with the UCSF Chimera package [27]. Functional protein domains were named according to the Conserved Domains Database (CDD, <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>).

2.2 Construction of plasmids

Total RNA was isolated with the RNeasyTM Plant Mini Kit (Qiagen) from Arabidopsis leaves and reverse transcribed to cDNA with the RevertAid H Minus First Strand cDNA Synthesis Kit (Fermentas). The AtNAA70 sequence was amplified from cDNA with specific primers containing appropriate restriction endonuclease sites (Supporting Information Table 1) using Phusion High fidelity DNA Polymerase (Finnzymes). The cDNAs encoding for full-length AtNAA70 was cloned in pFF19-EYFP, which allowed expression of AtNAA70 fused with EYFP at the C-terminus. The cDNA encoding for AtNAA70 lacking the predicted transit peptide was cloned in the pETM41 vector to allow expression of His₆-MBP-AtNAA70 in *E. coli*.

2.3 Subcellular localization of AtNAA70

Transformation of isolated protoplasts from rosette leaves of six-week-old *Arabidopsis thaliana* ecotype Col-0 was performed according to [28] using 10–20 μg of pFF19-AtNAA70:EYFP DNA. Protoplasts expressing AtNAA70:EYFP for two days were placed on a slice in a drop of 20 μl WI solution and imaged using Zeiss confocal

microscope LSM510META equipped with lasers for 405-, 488- and 543-nm excitation. Images were documented with 63x lens in multi-tracking mode. YFP fluorescence was excited at 514 nm and emission was recorded with a 560–615 nm band-pass filter. Chlorophyll auto-fluorescence was excited at 488 nm and emission was collected within 647–745 nm. Images were analyzed with the LSM 510 software suite (Zeiss).

2.4 Expression and purification of AtNAA70 in *E. coli*

E. coli Rosetta (DE3, Novagen) was transformed by electroporation with the pETM41-AtNAA70 plasmid encoding for a His₆-MBP-AtNAA70 fusion protein, and grown overnight on LB plate containing kanamycin. The His₆-MBP-AtNAA70 protein was expressed and purified by immobilized metal affinity chromatography as described in [29].

2.5 In vitro Lys-auto-acetylation assays

In vitro auto-acetylation assay was performed according to [30] with two independent methods. For determination of lysine acetylation with a N^ε-Lyse specific antiserum, 1 μg of purified His₆-MBP-AtNAA70 was incubated in a total volume of 80 μL reaction buffer (50 mM Tris-HCl (pH 8.5), 10% glycerol, 1 mM EDTA, 125 μM Acetyl-CoA). The degree of lysine acetylation was quantified (auto-Kat activity) by immunological detection using a specific anti-acLys antiserum (New England Biolabs, 1:10.000), the anti-Rabbit IgG-HRP antiserum (Promega; 1:20.000) and the SuperSignalTM West Dura Extended Duration Substrate (Thermo Scientific). Additionally, the auto-Kat activity of AtNAA70 was quantified by determination of incorporated [³H] label with a Tri-Carb 2810TR scintillation counter (PerkinElmer) after incubation of 20 μg purified His-MBP-AtNAA70 at 37°C in acetylation buffer (50 mM Tris-HCl (pH 8.5), 10% glycerol, 1 mM EDTA) containing 60 μM [³H]-acetyl-CoA (7.4 GBq/mmol, Hartmann Analytix). Free [³H]-acetyl-CoA was removed after incubation with a PD-SpinTrap (GE Healthcare) according to the manufacturer's protocol.

2.6 Sample preparation and N-terminus peptide enrichment

Bacterial cells over-expressing AtNAA70 were centrifuged at 5000 rpm 4°C and pellets frozen at -80°C. The cell pellets were resuspended in 50 mM Tris pH 8 plus 250 mM NaCl, subjected to sonication and cell debris were removed by centrifugation at 15 000 rpm for 30 min at 4°C. 1 mg of supernatant proteins was denatured and reduced before cysteine alkylation using iodoacetamide [19]. Then, to perform stable isotope labeling protein N-terminal acetylation quantitation (SILProNAQ), the sample was chemically treated to acetylate the free amino groups with d3-acetyl groups [31].

Rapidly, 25 μmol of N-acetoxy-[²H₃]-succinimide in DMSO per mg of protein was added to the alkylated proteins resuspended in 50 mM phosphate buffer at pH 7.5. After 90 min at 30°C, potential O-acetylation of Ser, Thr and Tyr side chains were reversed by adding 10 μl of 50% of hydroxylamine (v/v) and incubated for 20 min at room temperature. After a purification step using cold acetone precipitation, the protein pellet was resuspended in 50 mM ammonium bicarbonate and digested by the addition of 1/100 (w/w) of TPCK treated porcine trypsin (Sigma-Aldrich) for 1.5 h at 37°C twice. Peptides were desalted by Sep-PakTM solid phase extraction and the retained material was eluted with 80% ACN, 0.1% TFA followed by evaporation to dryness. The collected material was resuspended in strong cation exchange (SCX) LC buffer A (5 mM KH₂PO₄, 30% ACN and 0.05% formic acid) and injected into Summit LC system (Dionex, Sunnyvale, CA) equipped with Polysulfoethyl A 200 × 2.1 mm 5 μm 200 Å column (PolyLC, Columbia, MD). Peptides were eluted with a gradient of increasing KCl concentrations (SCX-LC Buffer B: 350 mM KCl in SCX-LC Buffer A; 0–5 min, 0% B; 15–40 min, 5–26% B; 40–45 min, 26–35% B). Fractions were collected every 2 min for 40 min and solvent was evaporated to dryness before storage at -20°C.

2.7 LC-MS/MS analysis

Analysis of the SCX-LC fractions was performed as previously described [32]. Rapidly, selected samples are loaded on the pre-column (NS-MP-10, Nano-separation,) of the Easy nLC II system (Thermo Scientific) followed by the separation on a C18 reverse phase column (NikkyoTechnos capillary column,) at a flow of 300 nL/min on 40 min gradient. The nano-LC was coupled to an OrbitrapTM Velos (Thermo Scientific). The survey scan was acquired by Fourier-Transform MS scanning 400–2000 Da at 30 000 resolution using internal calibration (lock mass) using the Top-20 acquisition method with 20 s exclusion time. Raw data files were extracted and exported with Proteome Discoverer (Thermo Scientific, Ver. 1.4) for ion signal higher than 1 counts and S/N higher than 1.5.

2.8 Data processing for protein identification and quantification

MASCOT 2.4 software was used for protein identification and co-/post-translational modification characterization using the *E. coli* K12 strain reference proteome subset extracted from UniProtKB (version 112) which also included the sequence of the recombinant proteins AtNAA10, AtNAA15 and AtNAA70. Trypsin/P rules were used with parent and fragment mass tolerance defined as 10 ppm and 0.7, respectively. Carbamidomethylcysteine and d3-acetyl on Lys were considered as fix modifications whereas Met-oxidation, protein NTA and d3-NTA were considered as variable modifications. All

data were filtered at 1% protein false discovery rate and only peptides with score higher than 25 are retained for the final data treatment. To extract specifically N-terminal peptides, MASCOT searches were exported in xml format and submitted to an in-house script written in Python. The parsing function searched for modifications and collected peptides with defined modifications such as d0/d3-NTA.

SILProNAQ quantification required the aid of MASCOT Distiller (ver. 2.5.1, Matrix Science) to combine the processing of the raw data and MASCOT identification results. MASCOT distiller extraction parameters were optimized for the Orbitrap “Top-20” acquisition files with a minimum S/N of 1, precursor charge from 1–5, Corr. Thr. > 0.7 and no grouping assignments. MASCOT distiller submissions for protein identification were performed using the MASCOT 2.4 identification tool against the reference proteome for *E. coli* K12 strain extracted from UniProtKB (version 112) using Trypsin/P cleavage rules. The parent and fragment mass tolerance were defined to 10 ppm and 0.7 Da, respectively, with carbamidomethylcysteine and d3-acetyl on Lys as fix modifications. Few variable modifications were used including oxidized Met (Met-Ox) and d0/d3-NTA quantitation method. Based on characterized proteins, MASCOT distiller determined NTA quantitation from the raw data. Since this tool was designed to deliver quantification values at the protein level (including all characterized peptides and not the protein N-terminus only), we developed an in-house Python script able to parse MASCOT distiller xml export files and to recalculate the NTA yield based only on the data related to the N-terminus peptides sharing a common starting position.

This tool extracts a few parameters including H/L ratio, signal quality coefficient (MASCOT distiller defined parameters: fraction, *E* value, correlation) and peptide related data (peptide start position, MASCOT score, peptide mass error). For each distinct N-terminus position, EnCOUTer aggregates the data passing the filtering criteria defined by the user and especially the signal quality coefficient (Corr > 0.5, *E* value < 0.1, score > 30, std. err. < 0.1, fraction > 0.5). Then, the geometric mean and geometric deviation was re-calculated including all data associated with the same N-terminus position that pass the filtering criteria defined by the user. H/D value is finally converted to NTA yield. Filtering parameters defined and provided by MASCOT distiller were also used in the extraction script, i.e. *E* val. < 0.1, std. err. < 0.1, MASCOT score > 30, corr. > 0.5, fraction > 0.5, Sc. *P* > 0.3. NTA yield was determined from the d0/d3 ratio and expressed in % of NTA for the different N-termini.

3 Results

3.1 Identification of potential chloroplastic N^α-acetyltransferases in *Arabidopsis thaliana*

Several studies demonstrated the presence of acetylated N-termini from plastid localized proteins [18–20].

However, the few characterized plant Nat proteins are all localized in the cytosol [19, 23], and thus could not be responsible for the N-terminal acetylation of these proteins.

In a search of plastidic Nats, sequences of known Nats from human, yeast and *Arabidopsis* were used as templates in a blast search on TAIR to find unknown Nat candidates [13, 14, 23, 33]. This sequence homology based search revealed 25 putative Nats in the *Arabidopsis* genome (Supporting Information Fig. 1). According to the SubCellular Proteomic Database (<http://suba.plantenergy.uwa.edu.au/>) seven of the 25 candidates are predicted to localize in chloroplast (listed in Supporting Information Table 2). The protein encoded by AT2G39000 displayed the highest conservation of the acetyl-CoA binding motif RxxGxG/A, which is conserved in all major Nats of eukaryotes [34]. Transcription of the AT2G39000 gene has been verified by qRT-PCR in roots, stem, flower, rosette and cauline leaves as according to [35] and was found to be approximately five-fold higher in rosette and cauline leaves when compared to non-green tissues (flower and root). However, AT2G39000 is ubiquitously transcribed during development of *Arabidopsis thaliana* according to public available transcriptome profiling repository databases (Supporting Information Fig. 2).

3.2 Tertiary structure modeling of potential chloroplastic N^α-acetyltransferases

We decided to use the structural information of HsNAA50 (2PSW) for a modeling approach based on pairwise alignment to gain more information about AT2G39000. After removal of the putative transit peptide domain (residues 1–55) a partial 3D model of AT2G39000 was obtained by modeling residues 63–112 (N-domain, ~40% similarity to 2PSW) and 195–254 (C-domain, ~45% similarity to 2PSW) to 2PSW, followed by superposition of both domains on 2PSW (Fig. 1A). The C-domain included the conserved NAT_SF (N-acetyltransferase family) domain that is known to bind the acetyl-CoA molecule by the RxxGxG/A motif (Fig. 1B). Two residues of this motif (R206 and R207) together with charged/polar residues such as K212, N238, K243 and D247 were recognized to contribute to the modeled acetyl-CoA binding pocket of AT2G39000 (Fig. 1B). The superimposed N-domain and C-domain of AT2G39000 resemble remarkably the tertiary structure of 2PSW (61% coverage) at the corresponding areas (Fig. 1A and D). Thus, we named AT2G39000 according to the revised Nat nomenclature system [13] AtNAA70. The N- and C-domain of AtNAA70 were separated by a linker region that is absent in HsNAA50 (Fig. 1C and D).

3.3 AtNAA70 is localized in chloroplasts

AtNAA70 is predicted by TargetP 1.1 server (<http://www.cbs.dtu.dk/services/TargetP/>) to target exclusively to the

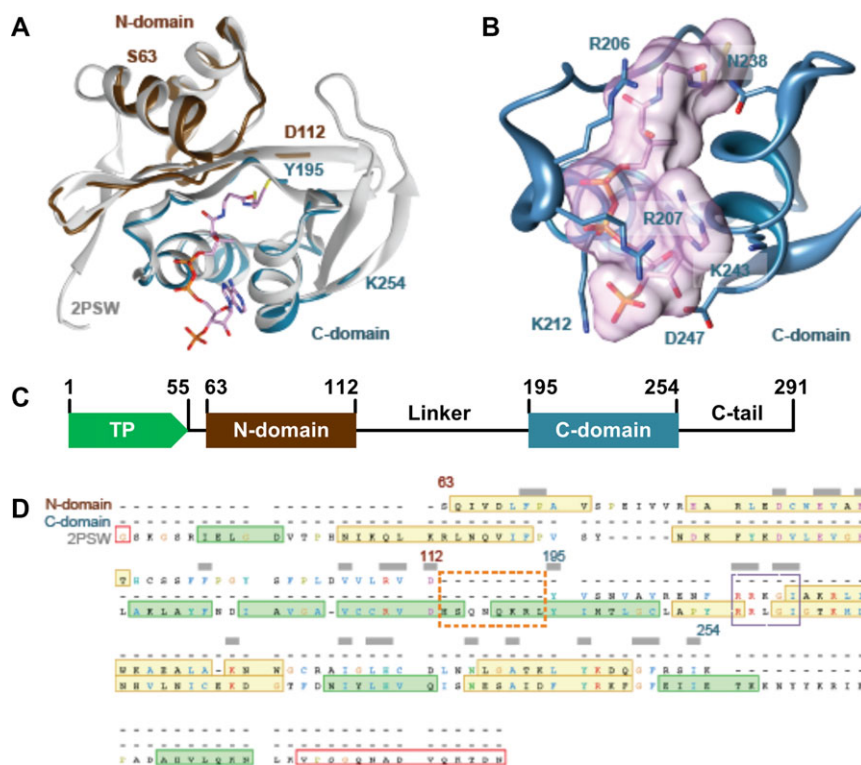


Figure 1. Structure modeling of chloroplastic AT2G39000 from *Arabidopsis thaliana*. (A) Superposition of modeled N-domain (brown) and C-domain (blue) from AT2G39000 with the template 2PSW (light gray) is shown. The enzyme structures are in cartoon representation, whereas CoA molecule is displayed in sticks representation. The modeled domains are labeled by the first and the last modeled residues. (B) Enlargement of modeled AT2G39000 C-domain harboring a CoA molecule (with its molecular surface). The CoA molecule and selected residues that probably interact with CoA are shown in stick presentation. (C) Schematic representation of AT2G39000 domain organization. TP, transit peptide. (D) Structural alignment of N- and C-domains with 2PSW. The orange box highlights the linker of AT2G39000 that is missing in 2PSW. The violet box designates the acetyl-CoA binding motif (RxxGxG/A). Conserved residues are marked by horizontal bars above the sequences. Green areas represent α -helices, yellow areas indicate β -strands.

chloroplast (TargetP scores: 0.928, Supporting Information Table 2). In order to confirm its subcellular localization, AtNAA70 fused to EYFP at the C-terminus was transiently expressed in *Arabidopsis* mesophyll protoplasts. The AtNAA70-EYFP signal perfectly overlaid with chloroplast auto-fluorescence, demonstrating its chloroplastic localization (Fig. 2). No YFP signal was detected in control protoplasts expressing the empty pFF19EYFP vector (Fig. 2).

3.4 AtNAA70 displays N^{acetyl}transferase activity toward *E. coli* proteins

AtNAA70 lacking the chloroplastic transit peptide (Supporting Information Table 2) in fusion with the 6x histidine-tag and the maltose binding protein (His₆-MBP) was expressed in *E. coli* cells, where acetylation is a very rare event [1, 11, 40]. Thus, the comparison of the N-terminal status of the wild type *E. coli* proteome with the proteome of cells expressing

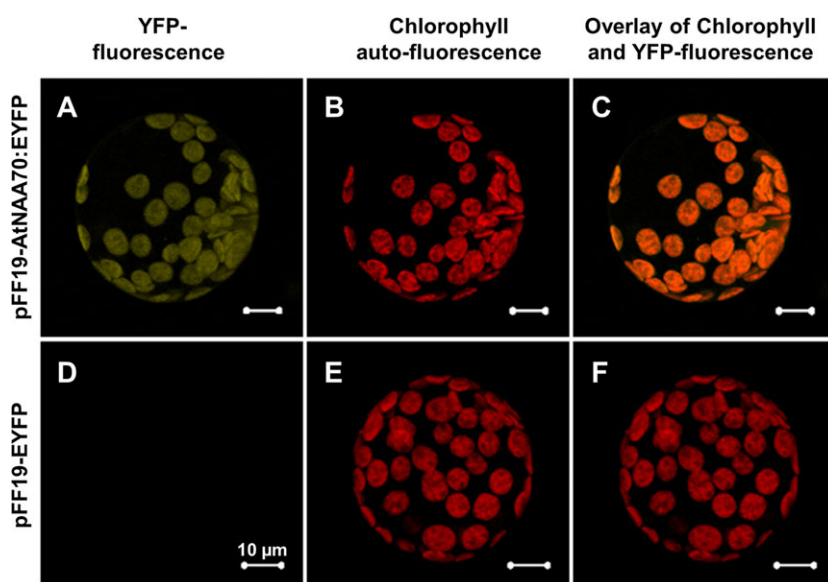


Figure 2. AtNAA70 is localized in chloroplasts. Full length AtNAA70 (AT2G39000) in fusion with EYFP was ectopically expressed under control of the 35S-promotor in mesophyll protoplasts extracted from six-week-old *Arabidopsis* leaves. The YFP signal (green) of AtNAA70-EYFP (A) perfectly overlap (C) with chlorophyll auto-fluorescence (red signal, B) determined in the same protoplasts. Protoplasts transiently transformed with the empty pFF19-EYFP vector (D–F) display chlorophyll auto-fluorescence (E) but no specific EYFP signal (D) using the same settings.

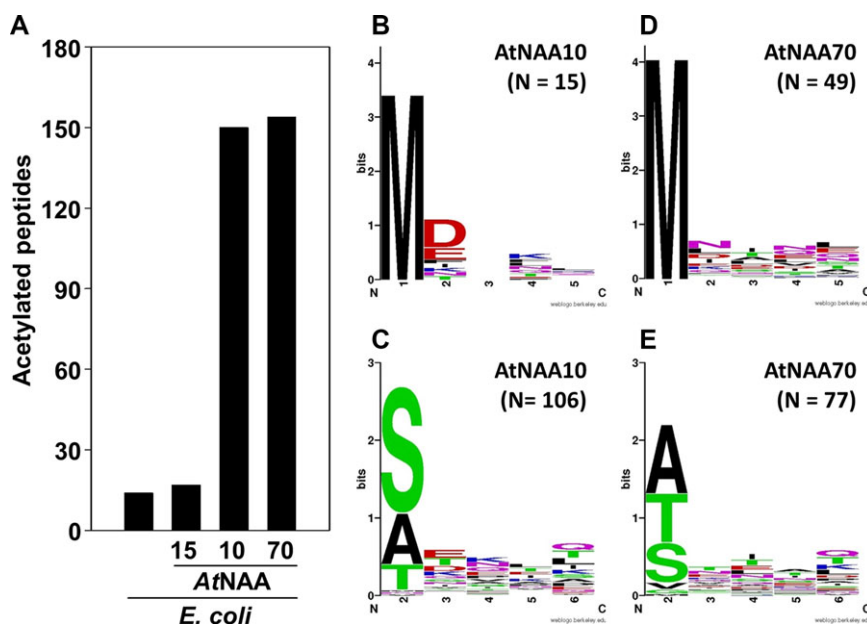


Figure 3. A newly established MS/MS-based in vivo Nat activity test demonstrates enzymatic acetylation activity of AtNAA70. (A) Number of identified acetylated N-terminal peptides after digestion of *E. coli* soluble protein lysates (black bar) with trypsin and enrichment of N-terminal peptides by SCX. Wild type *E. coli* and *E. coli* cells expressing either the auxiliary subunit (AtNAA15, negative control) or the catalytically active subunit of NatA (AtNAA10, positive control) were analyzed to verify the suitability of the new enzymatic Nat test. Expression of catalytic subunit of the NatA and the candidate Nat, AtNAA70, resulted in significant increase of acetylated peptides. (B–E) Web logos of N-termini from *E. coli* proteins that were found to be specifically acetylated after expression of AtNAA10 (B, C) or AtNAA70 (D, E). These proteins start with the iMet (B, D) or were subject of iMet removal (C, E). (N = number of acetylated peptides).

His₆-MBP-AtNAA70 or other plant NAAs should reveal N-termini of proteins acetylated by the recombinant plant proteins. As controls, the already characterized enzymatically active subunit, AtNAA10, and its auxiliary subunit, AtNAA15, of the plant NatA complex were also expressed in *E. coli* as His₆-MBP-tagged fusion proteins [Linster et al., in revision]. After SCX-LC enrichment, N-termini enriched fractions were analyzed by LC-MS/MS to determine the NTA frequency of the bacterial proteome induced by the expression of the plant Nat subunit. Using this large-scale proteomics approach more than 2000 peptides were identified, which correspond to more than 400 non-redundant protein N-termini in each sample (PRoteomics IDentifications database, <http://www.proteomexchange.org/databases/pride:PX001947>).

As expected only 17 acetylated peptides were identified in the proteome of cells expressing the His₆-MBP-AtNAA15 (negative control) of which 14 were acetylated also in wild type *E. coli* cells (Fig. 3A). In contrast, the proteome of cells expressing His₆-MBP-AtNAA10 comprised 150 acetylated N-termini of which 121 peptides were specifically found in cells expressing AtNAA10 (Supporting Information Table 3, Fig. 3A). The almost 10-fold increase in acetylation of protein N-termini demonstrated the suitability of the here proposed GAP approach for identification of substrate peptides. Only 15 of the 121 uniquely found acetylated peptides in the presence of AtNAA10 starts with an iMet that is predominantly followed by a negatively charged amino residue (Fig. 3B). The bulk of acetylated peptides correspond to peptides derived from proteins that were subjected to iMet removal (106, Fig. 3C), which matches the in vivo substrate specificity of other eukaryotic NAA10 [36, 37]. In an alternative in vitro methodology, a proteome-derived peptide library was used to characterize substrate specificity of purified HsNAA10 (NatA catalytic subunit). This approach evi-

denced a strong activity of the HsNAA10 against Glu-starting N-termini [37]. This NTA of Glu-starting N-termini by free HsNAA10 could not be observed for free AtNAA10 in the GAP approach. However, we could confirm the promoting function of Glu and Asp residues at position 2 for acetylation of N-termini by free AtNAA10 that was also demonstrated for HsNAA10 with the proteome-derived peptide library approach.

The expression of His₆-MBP-AtNAA70 also caused a significant increase in acetylation of N-termini of soluble *E. coli* proteins (Supporting Information Table 3, Fig. 3A). In the case of His₆-MBP-AtNAA70 126 specifically acetylated peptides were identified of which 49 start with iMet (Fig. 3D). The residual acetylated peptides (77) derived from N-termini that were subjected to iMet removal. The web logo for these peptides demonstrate a preference of NAA70 to acetylate N-termini starting with A > T > S (Fig. 3E). The following residue had only minor impact on the specificity of AtNAA70 using the here applied conditions. In order to verify the identity of AtNAA70 as a true catalytically active Nat, we purified the protein and confirmed its acetylase activity on protein N-termini according to the here determined substrate specificity (Supporting Information Fig. 3) with an accepted in vitro Nat activity test [36].

Since AtNAA70 should modify the mature N-termini of the imported and plastome-encoded proteins, it was interesting to compare the substrate specificity of AtNAA70 characterized in this study (Fig. 4A) with acetylated protein N-termini experimentally found in Arabidopsis chloroplasts (Fig. 4B and C). Remarkably, AtNAA70 accepts *E. coli* proteins as substrates that share the N-terminal amino acid with acetylated chloroplastic proteins (Fig. 4, arrow). This indicates that AtNAA70 activity contributes to the multitude of N-terminal acetylation events demonstrated in Arabidopsis chloroplasts.

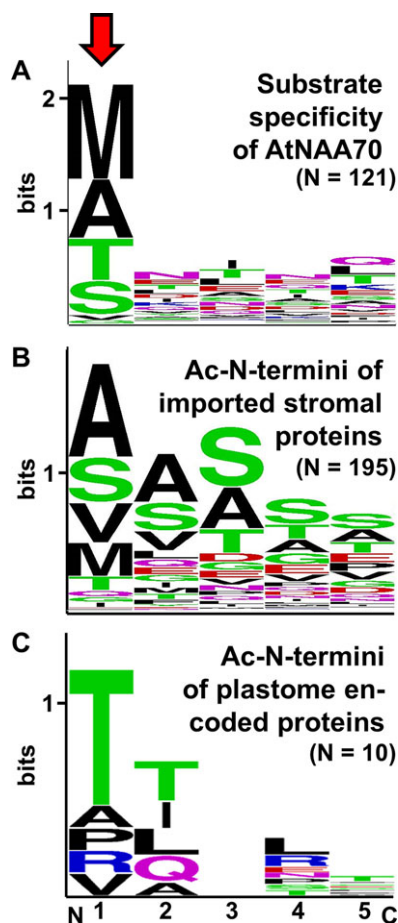


Figure 4. Comparison of AtNAA70 substrate specificity with N-termini of known acetylated plastid localized proteins. Web logo of substrate specificity for His₆-MBP-NAA70 ($N = 121$, A) is compared with experimentally identified acetylated N-termini of chloroplastic stromal proteins (B, C). The proteins ($N = 195$; [19]) are either imported from the cytosol (C) or translated ($N = 10$) in the chloroplast (D). The red arrow indicates similarity between substrates acetylated by AtNAA70 in *E. coli* and acetylated chloroplastic proteins in Arabidopsis.

3.5 AtNAA70 auto-acetylates residues K217, K254 and K265

Since HsNAA50 can acetylate the ϵ -amino group of Lys residues [30], we screened for additional N^ε-Lys-acetylated proteins after expression of His₆-MBP-AtNAA70 in *E. coli*. No significant increase in N^ε-Lys acetylation was found on *E. coli* proteins. However, three internal lysine residues (K217, K254 and K265) of His₆-MBP-AtNAA70 were evidenced to be acetylated (Fig. 5A and B). In order to test the possibility of N^ε-Lys auto-acetylation, purified His₆-MBP-AtNAA70 was incubated with or without acetyl-CoA for different periods of time. Indeed, samples incubated with acetyl-CoA displayed an increase of N^ε-acetylated Lys residues as determined with a N^ε-Lys specific antiserum in a time dependent manner,

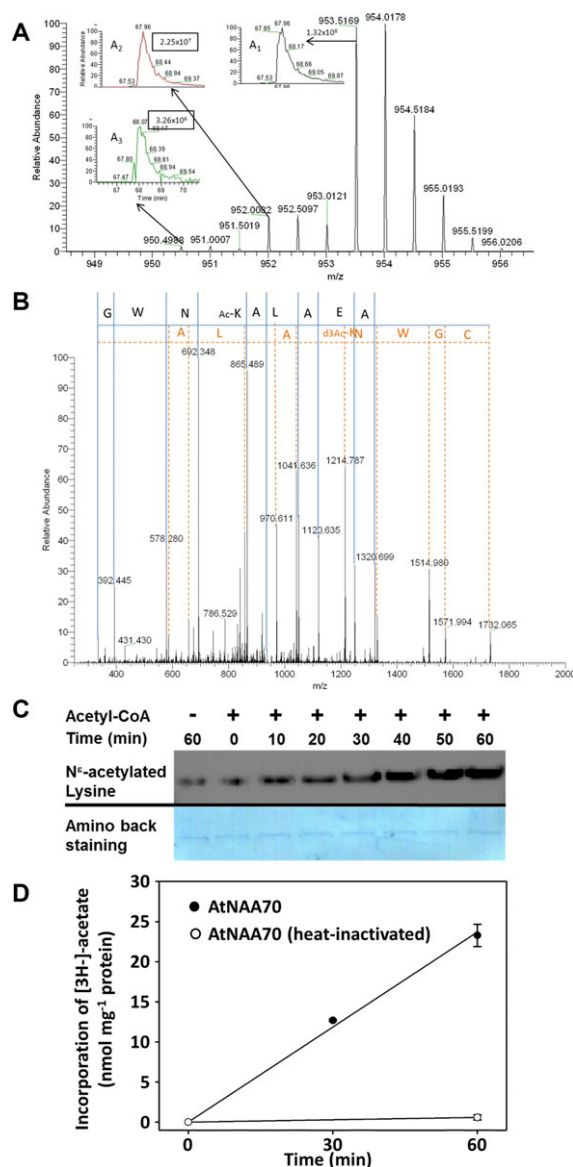


Figure 5. AtNAA70 possesses N^ε-acetylation activity on internal Lys residues. (A) MS spectrum of the AtNAA70 internal LIWKAEALAKNWGCR peptide with three distinct isotope distributions corresponding to the peptide with A₁) two heavy acetylations (d3-Acetyl, chemical modification), A₂) a mix of heavy and light acetylations (d3-Acetyl/Ac), A₃) two light acetylations (endogenous). The MS signal extractions displayed in A₁ to A₃ are used to determine the relative abundance of each form (unmodified peptide: 84%, partial Lys acetylation (on Lys 217 or 223): 14%, Acetylation of both Lys 217 and 233: 2%). (B) Annotated synthetic MSMS of the peptide (combination of spectra with precursor mass of 953.45 ± 0.5 Da). (C) Immunological detection of ac-Lys on AtNAA70 in absence (-) or presence (+) of acetyl-CoA for indicated time points using an N^ε-acLys specific antiserum. Staining of AtNAA70 with amido black serves as loading control. (D) Determination of AtNAA70 auto-Kat activity by incorporation of [³H]-acetyl moieties into purified AtNAA70 (black circles) over time. Incubation of [³H]-acetyl-CoA with heat inactivated AtNAA70 (white circles) for 60 min served as a negative control. ($N = 3$)

whereas no significant difference between the control sample at time 0 and the sample treated for 60 min without acetyl-CoA was observed (Fig. 5C). Signals detected at time point 0 confirm that Lys residues of His₆-MBP-AtNAA70 are partly acetylated in *E. coli* (Fig. 5A). To independently validate the N^ε-Lys auto-acetylation (auto-Kat) activity of AtNAA70, purified AtNAA70 was incorporated with ³H-acetyl-coenzyme A for 60 min. The time dependent incorporation of ³H-label into the purified proteins confirmed the determined auto-Kat activity by immunological detection of N^ε-Lys residues (Fig. 5D).

4 Discussion

4.1 Establishment of an in vivo Nat activity test by global acetylome profiling (GAP test)

Hitherto, substrate specificities of cytosolic Nats have been characterized by large-scale proteomics of respective loss-of-function mutants or by in vitro enzymatic tests using recombinant purified enzymes. Both approaches have specific advantages and drawbacks. Analysis of loss-of-function mutants is the gold-standard to assess endogenous substrates of Nats in vivo [14]. However this approach is not applicable, when the respective Nat is essential in higher eukaryotes or genomic analyses indicates high degree of functional redundancy of gene products [38, 39]. The determination of Nat substrate specificity in vitro is hampered by several obstacles: (i) the instability of the purified eukaryotic Nats ([30, 36], own observation), (ii) the necessity to provide large numbers of synthetic peptides as substrates [37] and (iii) the lack of high-throughput test systems.

Thus, we designed an in vivo Nat enzymatic activity test that is based on GAP test after recombinant expression of eukaryotic Nat subunits in a heterologous prokaryotic system. In a very similar approach, which was based on expression of HsNAA60 in yeast, the contribution of NatF to the evolutionary shift in N-terminal acetylation from yeast to humans has been analyzed [14]. Furthermore, the HsNAA40 catalytic subunit was biochemically characterized in the yeast *naa40* genetic background by a heterologous expression approach resulting in wild type like complementation [33]. Both approaches were successfully exploited to determine the substrate specificity of eukaryotic Nats, since the orthologous endogenous eukaryotic Nats were absent in the transfected yeast genotype. Like in these approaches the GAP test eliminates the necessity to purify the enzymatically active form of the acetyltransferase and provides a broad range of potential substrates for acetylation. The identification of false positive in vivo N-terminally acetylated substrates is minimized in the GAP test, since the prokaryotic machinery for NTA is limited to a low percentage of substrates of which only a handful were characterized with high NTA yield [1, 11, 40]. The validation of the GAP test by expressing the AtNAA10 and retrieving the same in vivo substrate specificity known for

eukaryotic NAA10 demonstrates its suitability [1, 37]. The use of the GAP test can be envisaged for numerous applications. In addition to substrate specificity determinations, kinetic information for the activity of candidate Nats can be obtained by precise pulse-chase experiments. The GAP test has the potential to test for acetyltransferase activity and substrate specificity of candidate proteins in low-throughput screens as a result of its simple design. It allows furthermore screening for potential auto-modifications of the candidate Nat as it has been demonstrated here for AtNAA70 by acetylation of three internal Lys residues. The acetylation of internal Lys has been independently confirmed by immunological detection of the N^ε-Lys acetylation activity (Kat) of AtNAA70. Kat activity has been also reported for HsNAA10 and HsNAA50 [30, 41] and regulates substrate specificity of HsNAA50 [30]. Finally, the GAP test can be applied in future studies to characterize the impact of potential modifiers of Nat substrate specificity, by simple co-expression of the eukaryotic Nat (e.g. AtNAA70) and the potential eukaryotic Nat modifier in *E. coli*. Prime candidates for these modifiers are auxiliary subunits of Nats [42] and ribosome-associated protein biogenesis factors, which are absent or significantly different in prokaryotes when compared to eukaryotes [2].

4.2 Structural conservation and substrate specificity of AtNAA70

The in silico analysis of the Arabidopsis genome revealed seven potential candidates for chloroplast localized Nats. We experimentally confirmed the chloroplastic subcellular localization of one candidate, AtNAA70, unambiguously demonstrated its enzymatic Nat activity and revealed a remarkable structural conservation with a co-translationally acting cytosolic HsNAA50 [2, 30]. The tertiary structure of HsNAA50 was re-assembled by two distinct domains (N- and C-domain) of AtNAA70, which are separated by a linker. The assembly of functional structures by modular acting domains is an emerging concept in protein chemistry and contributes significantly to evolutionary driven construction of new protein functions [43]. The C-domain of AtNAA70 harbors the typical GNAT5 acetyl-CoA binding pocket, whose core is formed by a highly conserved RxxGxG/A domain [34].

AtNAA70 has a broad substrate specificity that partly overlaps with all types of known cytosolic Nat complexes (NatA to NatF, [1, 2]). The cytosolic Nat complexes are classified by their preference to accept substrates either starting with iMet or with the penultimate residue exposed by iMet removal (non-Met peptides) [1]. HsNAA10 bound in the NatA complex accepts almost solely non-Met peptides. Recently, the interaction of HsNAA15 with HsNAA10 in NatA has been demonstrated to allosterically reconfigure the HsNAA10 active site for sequence-specific N-terminal acetylation [42], which provides a molecular explanation for the fact that free HsNAA10 [37] and free AtNAA10 (this study) are able to acetylate substrates starting with iMet followed by acidic residues (Glu or

Asp). Nevertheless, even free NAA10 subunits of both species still prefer to acetylate non-Met peptides ([37], Fig. 3B and C). The usage of proteome-derived peptide libraries demonstrated that free HsNAA10 prefers to acetylate glutamic acid at position 1 (82% of all acetylated peptides [37]). Such a substrate preference was not observed for AtNAA10 by using the GAP test, most probably because endogenous proteins starting with Glu are rare in *E. coli*, since Glu at position two inhibits iMet removal in eukaryotic and prokaryotic systems [44].

In contrast, AtNAA70 is able to acetylate iMet and non-Met starting substrates with similar preference in the GAP test (Fig. 3D and E). Interestingly, the sole Nat ortholog in the archaea *Sulfolobus solfataricus* (SsNat) is also able to accept both types of substrates, because its active site represents a hybrid of the NatA and the NatE active sites [45]. Since SsNat is believed to represent an ancestral Nat from which the eukaryotic Nat machinery evolved, substrate specification of cytosolic Nat might be a secondary adapted feature during evolution of eukaryotes [45].

As a result of its broad substrate specificity AtNAA70 might be able to acetylate many of N-termini which were experimentally characterized to be acetylated in the chloroplast stroma (Fig. 4B and C). However, a minor number of Ile, Glu and Gln starting N-termini might not be acetylated by AtNAA70 according to the here determined substrate specificity (Fig. 4A), suggesting the existence of other chloroplastic Nats. Prime candidates for Nats targeting N-termini of proteins starting with Ile, Glu and Gln are the six putatively chloroplast localized Nat-like proteins identified in our homology based bioinformatics screen.

The MS proteomics data in this paper have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [46]: dataset identifier PXD001947. We thank the French National Research Agency (grant no. ANR-13-BSV6-0004) and the German Research Society (DFG grant no. He 1848/14-1, He1848/15-1, and SFB 1036/TP13) for support of W.V.B. and E.L., R.H. and M.W., respectively. T.V.D. was funded by the Ministry of Education and Training of Vietnam and the Schmeil Foundation, Heidelberg. This study has benefited from the facilities and expertise of the SICaPS platform of I2BC (Institute for Integrative Biology of the Cell).

The authors have declared no conflict of interest.

5 References

- Arnesen, T., Towards a functional understanding of protein N-terminal acetylation. *PLoS Biol.* 2011, 9, e1001074.
- Giglione, C., Fieulaine, S., Meinel, T., N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie* 2014. doi: 10.1016/j.biochi.2014.11.008.
- Behnia, R., Panic, B., Whyte, J. R., Munro, S., Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat. Cell. Biol.* 2004, 6, 405–413.
- Singer, J. M., Shaw, J. M., Mdm20 protein functions with Nat3 protein to acetylate Tpm1 protein and regulate tropomyosin-actin interactions in budding yeast. *Proc. Natl. Acad. Sci. U S A* 2003, 100, 7644–7649.
- Hershko, A., Heller, H., Eytan, E., Kaklij, G., Rose, I. A., Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad. Sci. U S A* 1984, 81, 7021–7025.
- Hwang, C. S., Shemorry, A., Varshavsky, A., N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 2010, 327, 973–977.
- Holmes, W. M., Mannakee, B. K., Gutenkunst, R. N., Serio, T. R., Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nat. Commun.* 2014, 5, 4383.
- Bradshaw, R. A., Brickey, W. W., Walker, K. W., N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends Biochem. Sci.* 1998, 23, 263–267.
- Driessen, H. P., de Jong, W. W., Tesser, G. I., Bloemendal, H., The mechanism of N-terminal acetylation of proteins. *CRC Crit. Rev. Biochem.* 1985, 18, 281–325.
- Gautschi, M., Just, S., Mun, A., Ross, S. et al., The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. *Mol. Cell. Biol.* 2003, 23, 7403–7414.
- Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., Sherman, F., Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J.* 1999, 18, 6155–6168.
- Polevoda, B., Hoskins, J., Sherman, F., Properties of Nat4, an N(alpha)-acetyltransferase of *Saccharomyces cerevisiae* that modifies N termini of histones H2A and H4. *Mol. Cell. Biol.* 2009, 29, 2913–2924.
- Polevoda, B., Arnesen, T., Sherman, F., A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. *BMC Proc.* 2009, 3, S2.
- Van Damme, P., Hole, K., Pimenta-Marques, A., Helsen, K. et al., NatF contributes to an evolutionary shift in protein n-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet.* 2011, 7, e1002169.
- Giglione, C., Boularot, A., Meinel, T., Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* 2004, 61, 1455–1474.
- Giglione, C., Meinel, T., Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci.* 2001, 6, 566–572.
- Kleffmann, T., von Zychlinski, A., Russenberger, D., Hirsch-Hoffmann, M. et al., Proteome dynamics during plastid differentiation in rice. *Plant Physiol.* 2007, 143, 912–923.
- Zybailov, B., Rutschow, H., Friso, G., Rudella, A. et al., Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 2008, 3, e1994.
- Bienvenut, W. V., Sumpton, D., Martinez, A., Lilla, S. et al., Comparative large scale characterization of plant versus

- mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol. Cell. Proteomics* 2012, 11, M111015131.
- [20] Wirtz, M., Heeg, C., Samami, A. A., Ruppert, T., Hell, R., Enzymes of cysteine synthesis show extensive and conserved modifications patterns that include N^ε-terminal acetylation. *Amino Acids* 2010, 39, 1077–1086.
- [21] Hoshiyasu, S., Kohzuma, K., Yoshida, K., Fujiwara, M. et al., Potential involvement of N-terminal acetylation in the quantitative regulation of the epsilon subunit of chloroplast ATP synthase under drought stress. *Biosci. Biotechnol. Biochem.* 2013, 77, 998–1007.
- [22] Bienvenut, W. V., Espagne, C., Martinez, A., Majeran, W. et al., Dynamics of post-translational modifications and protein stability in the stroma of *Chlamydomonas reinhardtii* chloroplasts. *Proteomics* 2011, 11, 1734–1750.
- [23] Pesaresi, P., Gardner, N. A., Masiero, S., Dietzmann, A. et al., Cytoplasmic N-terminal protein acetylation is required for efficient photosynthesis in Arabidopsis. *Plant Cell* 2003, 15, 1817–1832.
- [24] Ferrandez-Ayela, A., Micol-Ponce, R., Sanchez-García, A. B., Alonso-Peral, M. M. et al., Mutation of an Arabidopsis NatB N-alpha-terminal acetylation complex component causes pleiotropic developmental defects. *PLoS ONE* 2013, 8, e80697.
- [25] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J. et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997, 25, 3389–3402.
- [26] Šali, A., Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993, 234, 779–815.
- [27] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S. et al., UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004, 25, 131605–131612.
- [28] Yoo, S. D., Cho, Y. H., Sheen, J., Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2007, 2, 1565–1572.
- [29] Wirtz, M., Berkowitz, O., Droux, M., Hell, R., The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and protein-protein interaction. *Eur. J. Biochem.* 2001, 268, 686–693.
- [30] Evjenth, R., Hole, K., Karlsen, O. A., Ziegler, M. et al., Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity. *J. Biol. Chem.* 2009, 284, 31122–31129.
- [31] Chakraborty, A., Regnier, F. E., Global internal standard technology for comparative proteomics. *J. Chromatogr. A* 2002, 949, 173–184.
- [32] Bienvenut, W. V., Sumpton, D., Lilla, S., Martinez, A. et al., Influence of various endogenous and artefact modifications on large scale proteomics analysis. *Rapid Commun. Mass Spectrom.* 2013, 27, 443–450.
- [33] Hole, K., Van Damme, P., Dalva, M., Aksnes, H. et al., The human N-alpha-acetyltransferase 40 (hNaa40p/hNatD) is conserved from yeast and N-terminally acetylates histones H2A and H4. *PLoS One* 2011, 6, e24713.
- [34] Zhu, H. Y., Li, C. M., Wang, L. F., Bai, H. et al., In silico identification and characterization of n-terminal acetyltransferase genes of Poplar (*Populus trichocarpa*). *Int. J. Mol. Sci.* 2014, 15, 1852–1864.
- [35] Haas, F. H., Heeg, C., Queiroz, R., Bauer, A. et al., Mitochondrial serine acetyltransferase functions as a pacemaker of cysteine synthesis in plant cells. *Plant Physiol.* 2008, 148, 1055–1067.
- [36] Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M. et al., Identification and characterization of the human ARD1-NATH protein acetyltransferase complex. *Biochem. J.* 2005, 386, 433–443.
- [37] Van Damme, P., Evjenth, R., Foyn, H., Demeyer, K. et al., 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. *Mol. Cell. Proteomics: MCP* 2011, 10, M110004580.
- [38] Rope, Alan F., Wang, K., Evjenth, R., Xing, J. et al., Using VAAST to Identify an X-linked disorder resulting in lethality in male infants due to n-terminal acetyltransferase deficiency. *Am. J. Human Genet.* 2011, 89, 28–43.
- [39] Wang, Y., Mijares, M., Gall, M. D., Turan, T. et al., Drosophila variable nurse cells encodes arrest defective 1 (ARD1), the catalytic subunit of the major N-terminal acetyltransferase complex. *Dev. Dyn.* 2010, 239, 2813–2827.
- [40] Soppa, J., Protein acetylation in archaea, bacteria, and eukaryotes. *Archaea* 2010, 1–9.
- [41] Jeong, J. W., Bae, M. K., Ahn, M. Y., Kim, S. H. et al., Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation. *Cell* 2002, 111, 709–720.
- [42] Liszczak, G., Goldberg, J. M., Foyn, H., Petersson, E. J. et al., Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nat. Struct. Mol. Biol.* 2013, 20, 1098–1105.
- [43] Lam, B. C., Blumwald, E., Domains as functional building blocks of plant proteins. *Trends Plant Sci.* 2002, 7, 544–549.
- [44] Giglione, C., Vallon, O., Meinel, T., Control of protein lifespan by N-terminal methionine excision. *EMBO J.* 2003, 22, 13–23.
- [45] Liszczak, G., Marmorstein, R., Implications for the evolution of eukaryotic amino-terminal acetyltransferase (NAT) enzymes from the structure of an archaeal ortholog. *Proc. Natl. Acad. Sci. U S A* 2013, 110, 14652–14657.
- [46] Vizcaíno, J. A., Côté, R. G., Csordas, A., Dianes, J. A. et al., The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, 41, D1063–D1069.