



**HAL**  
open science

# The Global Acetylation Profiling Pipeline for Quick Assessment of Protein N-Acetyltransferase Specificity In Cellulo

Thierry Meinnel, Jean-Baptiste Boyer, Carmela Giglione

► **To cite this version:**

Thierry Meinnel, Jean-Baptiste Boyer, Carmela Giglione. The Global Acetylation Profiling Pipeline for Quick Assessment of Protein N-Acetyltransferase Specificity In Cellulo. *Methods in Molecular Biology*, 2023, Mass Spectrometry-Based Proteomics, 2718, pp.137-150. 10.1007/978-1-0716-3457-8\_8. hal-04235025

**HAL Id: hal-04235025**

**<https://hal.science/hal-04235025>**

Submitted on 10 Oct 2023

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

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

# **The Global Acetylation Profiling pipeline for quick assessment of protein N-acetyltransferase specificity in cellulose**

Thierry Meinzel<sup>1,iD,\*</sup>, Jean-Baptiste Boyer<sup>1</sup>, and Carmela Giglione<sup>1,iD,\*</sup>

<sup>1</sup> Université Paris Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette cedex, France

<sup>iD</sup> **ORCID:** 0002-7475-1558 (C. Giglione); 0001-5642-8637 (T. Meinzel)

\* **Correspondence:** Thierry Meinzel, [thierry.meinzel@i2bc.paris-saclay.fr](mailto:thierry.meinzel@i2bc.paris-saclay.fr) or Carmela Giglione, [carmela.giglione@i2bc.paris-saclay.fr](mailto:carmela.giglione@i2bc.paris-saclay.fr)

**Keywords:** N-acetyltransferase, acetylation, GNAT, N-terminomics, *Escherichia coli*, substrate specificity, heterologous expression, protein fusion, quantitative acetylomics.

**Running head:** Global Acetylation Profiling (GAP)

## **ABSTRACT**

Global Acetylation Profiling (GAP) consists of heterologous expression of a given N-acetyltransferase (NAT) in *Escherichia coli* to assess its specificity. The remarkable sensitivity and robustness of the GAP pipeline relies on the very low frequency of known N-terminal acetylated proteins in *E. coli*, including their degree of N-terminal acetylation. Using the SILProNAQ mass spectrometry strategy on bacterial protein extracts, GAP permits easy acquisition of both qualitative and quantitative data to decipher the impact of any putative NAT of interest on the N-termini of newly acetylated proteins. This strategy allows rapid determination of the substrate specificity of any NAT.

## 1 INTRODUCTION

To challenge the N-terminal specificity of an N-acetyltransferase (NAT), a number of approaches may be applied. N-terminomics methods were set up in the early 2000's to this aim and allow both qualitative and quantitative determination of the impact of a given NAT [1,2]. Usually, such methods allow for a differential analysis as they involve, when available, mutants devoid of a given NAT and comparison with the wild-type or even the mutant expressing a heterologous NAT. This latter approach was performed in the yeast *Saccharomyces cerevisiae* to characterize the human NatA and NatB orthologs [3,4]. Among eukaryotes, yeast is unique given that each of its NATs is not essential for cell growth as mutants devoid of one NAT subunit are viable under defined growth conditions [5,6]. These approaches may be challenging due to the poor signal to background noise ratio available when the specificity of the NAT is sharp or closely related to the baseline N-acetylation. Other large-scale analyses using N-terminomics strategies only indirectly provide information on the specificity of each NAT [7]. Although at a lower throughput, *in vitro* strategies with purified NATs and peptide libraries were also used to the characterization or to refine NAT substrate specificity [8].

Global Acetylation Profiling (GAP) was set to get rapid insight on whether or not a NAT candidate discloses a NAT activity and, if so, to which extent. It takes advantage of the unique features of *Escherichia coli* concerning N-terminal protein modifications. This organism displays indeed most of its N-termini devoid of the N-terminal formyl group inserted in the frame of protein synthesis and originally attached to the amino group of the first methionine [9]. This is achieved due to the massive proteolytic and co-translational action of peptide deformylases [10-13]. In addition, very few (less than 40), and well-known proteins undergo NTA and the level of NTA for the majority of proteins is low [9,14,15]. These few bacterial acetylation sites result from a couple of very selective NATs, known as RimI, RimJ and RimL with RimJ, behaving as the major NTA catalysts in *E. coli* [15,16]. The GAP strategy consists first in the expression of the NAT candidate of interest as a Maltose Binding Protein (MBP)-NAT fusion in *E. coli*. Second, a dedicated version of the SILProNAQ pipeline – a mass spectrometry strategy dedicated to

proteomic quantitative analysis of protein N-termini [17] – is applied. The GAP pipeline includes therefore robust N-terminomics strategies previously validated [17,18]. The pipeline was successfully applied to a number of *A. thaliana* NATs (reviewed in [19]). These NATs corresponds to NAA10, NAA50, NAA60, the three NAA70 and the five NAA90s [14,20-23]. The GAP with few adaptations has been used also for assessing lysine acetyltransferase (KAT) specificity (see protocol in [20]).

## **2 MATERIALS**

### **2.1 Materials**

1. SpeedVac Concentrator (Savant ISS110, Thermo Scientific) equipped with condensation trap and vacuum pump (vacuum < 1 Torr).
2. High Performance Liquid Chromatography (HPLC) UPHLC system with HS diode array detector, autosampler and thermostatted HS column compartment (Dionex UltiMate 3000, Thermo Scientific).
3. LTQ-Orbitrap Velos™ (Thermo Scientific) or equivalent ESI mass spectrometer with high survey-scan and resolution.
4. nano-Liquid Chromatography (n-LC) system (Easy nLC-II, Thermo Scientific).
5. Refrigerated microfuge (20R, Beckman Coulter) with Eppendorf tube rotor.
6. PolySULFOETHYL A (200 x 2.1 mm) column (5 µm particle size, 200 Å pore size) for Strong Cation-exChange (SCX) chromatography (Hichrom, UK).
7. Reverse phase nano-precolum (NanoSeparation, Netherlands).
8. Reverse phase analytical nano-column (Nikkyo Technos Co., Tokyo, Japan).

### **2.2 Main chemicals**

1. N-hydroxysuccinimide (purum, >97%., 56480, Sigma-Aldrich).
2. Acetic anhydride-d<sub>6</sub> (99 atom % D, 175641, Sigma-Aldrich).

3. Trifluoroacetic acid (TFA, 299537, Sigma-Aldrich).
4. Acetonitrile (34851-1L-M, Sigma-Aldrich).
5. Formic acid (98% pure, ACS reagent, 33015-1L-M, Sigma-Aldrich).
6. n-hexane, anhydrous, 95% (296090, Sigma-Aldrich).

### 2.3 Buffers and Solutions

7. Protein extraction buffer: 50 mM Hepes (H4034, Sigma-Aldrich)/NaOH pH 7.2, 1.5 mM MgCl<sub>2</sub>, (M2679, Sigma-Aldrich), 1 mM EGTA (E3889, Sigma-Aldrich); 10% glycerol (VWR, 24388.260), 1% Triton X-100 (T9284, Sigma-Aldrich), 0.15 M NaCl (71380, Sigma-Aldrich), 2 mM PMSF (P7626 Sigma-Aldrich), one tablet of protease inhibitor cocktail (EDTA+; 058892988001, Roche).
8. Denaturation solution: 6 M guanidine hydrochloride (Euromedex, EU0045), 50 mM Tris-Hcl pH 8, 4 mM dithiothreitol (Euromedex, EU0006).
9. Iodoacetamide stock solution (BIO BASIC, Ref: BI-IB0539): 0.55 M iodoacetamide in water, stored at -20 °C.
10. Phosphate buffer: 50 mM KH<sub>2</sub>PO<sub>4</sub> (P9791, Sigma-Aldrich) pH 7.5, adjusted with KOH.
11. N-acetylation reagent solution: 0.32 g of N-acetoxy-[<sup>2</sup>H<sub>3</sub>]-succinimide solubilised in 1 mL of dimethylsulfoxide (final concentration 2 M).
12. Digestion buffer: 50 mM ammonium bicarbonate (09830, Sigma-Aldrich) in water.
13. Trypsin stock solution: tosylphenylalanyl chloromethyl ketone (TPCK)-treated bovine trypsin (Sigma-Aldrich, T1426) solubilised in 1 mM HCl at a 10 mg/mL concentration (stored -20 °C).
14. Sep-Pak equilibration solution: 0.1% TFA in water.
15. Sep-Pak elution solution: 80% acetonitrile, 0.1% TFA, in water.
16. SCX buffer A: 5 mM KH<sub>2</sub>PO<sub>4</sub>, 30% acetonitrile, 0.05% formic acid, in water.
17. SCX buffer B: 0.35 M KCl in SCX buffer A.
18. LC-MS/MS loading buffer: 2% acetonitrile, 0.1% TFA, in water.

19. LC-MS/MS elution solvent A: 2% acetonitrile, 0.1% formic acid, in water.
20. LC-MS/MS elution solvent B: 0.1% formic acid in acetonitrile.
21. Luria-Bertani LB (Luria-Bertani) liquid medium; 950 mL H<sub>2</sub>O, 10g tryptone, 10 g NaCl, 5 g yeast extract.
22. 50 mg/mL kanamycin stock solution in H<sub>2</sub>O; sterilized by passing through a 0.2 µm syringe filter.
23. 1 M isopropyl-beta-D-thiogalactoside (IPTG) stock solution; sterilize through a 0.2 µm syringe filter.

## 2.4 NAT expression plasmids

1. Plasmids used for the GAP assay were pETM41 (Fig. 1) as the control and the pETM41-NAT vector expressing the NAT of interest. pETM41-AtNAA15 can be used alternatively for control induction as NAA15 does not have any N-acetyl transferase activity [14,23].
2. pETM41 (*see Note 1*) is the plasmid used for any NAT expression as a His<sub>6</sub>-MBP-NAT fusion (*see Notes 2-4*).
3. MBP-NAT fusions can be easily obtained by cloning in frame between the *NcoI* and any of the four *KpnI/BamHI/EcoRI/NotI* unique restriction sites of the polylinker region (*see Note 5*).
4. For expression, *E. coli* BL21(DE3)pLysS (Thermo Fisher) or *E. coli* Rosetta TM(DE3) cells (Merck) are transformed with the pETM41 derivative of interest.

[INSERT FIGURE 1 HERE]

## 3 METHODS

### 3.1 Preparation of N-acetoxy-[<sup>2</sup>H<sub>3</sub>]-succinimide

1. Solubilize 1.865 g of N-hydroxysuccinimide in 5 g of acetic anhydride-d<sub>6</sub> in a 15 mL Falcon tube (*see Note 6*).

2. Place the tube at gentle agitation (200 rpm) at room temperature and leave at least 15 h.
3. Collect crystals and remove excess solvent using filter paper (Whatmann, GE Healthcare).
4. Wash crystals twice using 200  $\mu$ L of anhydrous hexane, before being vacuum dried for a few minutes in 2mL Eppendorf tubes (*see Note 7*).

### **3.2 *E. coli* cell Extract Preparation**

1. Prepare 200 mL Erlenmeyer sterile containers filled with 30 mL of Luria-Bertani (LB) medium supplemented with kanamycin at 50  $\mu$ g/mL final concentration.
2. Dilute a fresh overnight bacterial pre-culture to start growth at 0.025 OD<sub>600</sub> in LB growth medium preheated at 37 °C.
3. Grow cultures at 37 °C until OD<sub>600</sub> reaches 0.2.
4. Add 0.5 mM IPTG.
5. Grow cells for another two generations in the exponential phase until OD<sub>600</sub>=0.9 to 1.0; i.e., approximately 1 h at 37 °C (*see Note 8*).
6. Cool cells on a mixture of ice and water at 4 °C
7. Briefly centrifuge cells at 5,250 g in a Falcon Tube (*see Note 9*).

### **3.3 Protein Extraction**

1. Add 300-500  $\mu$ L of protein extraction buffer to the cell pellet for solubilisation.
2. Lyse samples by ultrasonication, using the following parameters: 50% amplitude, 1 s ON pulse, 1 s OFF pulse, total time 30 s.
3. Place samples at 4 °C under agitation for 1 h to ensure proper protein extraction.
4. Centrifuge samples at 17,000 x g at 4 °C for 30 min.
5. Collect supernatants in 2 mL plastic microtubes, and directly use or store at -80 °C.
6. Check NAT overexpression at this stage (*see Note 10*) or later on (*see Note 11*).

### **3.4 Protein Denaturation and Labelling**



1. Transfer the relevant volume of protein extracts into 2 mL tubes (*see Note 12*).
2. Add cold acetone in at least 4 times the volume of the sample.
3. Vortex the samples and put them at -20 °C for at least 2 h.
4. Centrifuge samples at 17,000 g at -9 °C (or at the lowest available temperature, not more than 4 °C), for 30 min.
5. Discard supernatants and solubilize the pellets in 200 µL of denaturation solution.
6. Place the samples for 15 min at 95 °C to allow for protein denaturation.
7. Cool samples to room temperature.
8. Add 20 µL of iodoacetamide stock solution are mixed for cysteine alkylation.
9. Incubate the samples at room temperature and in the dark for 1 h.
10. Add 900 µL of cold acetone per sample.
11. Vortex the samples and place them -20 °C for at least 2 h.
12. Centrifuge the samples at 17,000 g at -9 °C for 30 min.
13. Discards the supernatants and solubilize the pellets in 200 µL of phosphate buffer.
14. Add 15 µL of 2 M N-acetylation reagent solution to each sample.
15. Incubate the samples at 30 °C under shaking (550 rpm) for 90 min.
16. Stop the reaction by adding of 10 µL of 50% hydroxylamine.
17. Agitate samples at room temperature at 550 rpm for 20 min.
18. Add 900 µl of cold acetone per sample. Vortex the samples and placed them at -20 °C for at least 2 h.

### **3.5 Protein digestion and desalting step**

1. Centrifuge samples at 17,000 g at -9 °C (or the lowest available temperature, not more than 4 °C), for 30 min.
2. Discard supernatants and solubilize the pellets in 200 µL of digestion buffer.
3. Add 1 µL of trypsin stock solution to each sample.

4. Incubate the samples at 37 °C and 550 rpm for 90 min.
5. Add another 1 µL of trypsin stock solution and repeat the incubation step.
6. Stop trypsin digestion by adding 1 mL of 1% TFA. At this stage, the peptide samples may be stored at -20 °C.
7. Recondition SepPak C18 cartridges with 3 mL of SepPak elution solution.
8. Equilibrate the SepPak cartridges using 10 mL of SepPak equilibration solution.
9. Load individual peptide mixtures on separate SepPak C18 cartridges and wash with 10 mL of SepPak equilibration solution.
10. Elute the retained and desalted peptides with 3 mL of SepPak elution solution.
11. Vacuum dry 1 mL of eluted peptides separately in 2 mL tubes.

### **3.6 Strong Cation Exchange (SCX) Peptide Fractionation**

1. Resuspend dried peptide samples in 100 µL of SCX buffer A.
2. Load 50 µL of peptide sample.
3. Fractionate peptides at a constant flow rate of 0.2 ml/min by HPLC with a SCX column and the gradient shown in Table 1.
4. Collect fractions from t=2 to t=30 min, switching tubes every 2 min, thus yielding 400 µL fractions.
5. Vacuum dry fractions 2-11 and stored at -20 °C.

### **3.7 LC-MS/MS Analysis**

1. Resuspend SCX fractions 2-11 in 30 µL of loading buffer.
2. Inject 5 µL per sample for LC-MS/MS analysis.
3. Load peptides onto a C18 nano-precolum, elute peptides and separate them on a reverse phase analytical nano-column, at a constant flow rate of 300 nL/min and using the gradient shown in Table 2. To several types of mass spectrometers can be used, we perform such analysis using an

ESI mass spectrometry instrument, starting with a MS survey scan acquired in the Orbitrap at 60,000 full width at half-maximum (FWHM) resolution, for a m/z range of 400-2000 with lock-mass based internal calibration, followed by the fragmentation and MS/MS analysis of the 20 most intense precursor ions. All precursor, even singly charged ions are considered for triggering MS/MS acquisitions.

### 3.8 Data Processing

1. Process the 10 raw files that are acquired for a single sample together in a Mascot Distiller project, using peak extraction parameters optimized for Orbitrap top-20 acquisition files.
2. Search the data against the *E. coli* protein database (<https://www.uniprot.org/proteomes/UP000000625>; see **Note 13**), with the following parameters: parent and fragment mass tolerance of 10 ppm and 0.5 Da respectively, a “semi-trypsin/P” enzyme digestion pattern, oxidized methionine as a variable modification, and carbamidomethyl-cysteine and d<sub>3</sub>-acetyl-lysine as fixed modifications. Include the N-terminal natural and labelled acetylations in the “Protein Nterm Acetylation” quantification method used (see **Note 10**).
3. After protein identification, perform quantification on all hits with a peptide score of at least 25 and a significance threshold of 0.2.
4. Export quantitation results from Mascot Distiller as a complete xml report. Hereby, each Mascot search result, corresponding to each individual fraction, is exported as an xml file from the server (see Figs. 2-3 for details).
5. Parse the Mascot Distiller and Mascot search outputs together using the eNcounter script [18] to validate potential N-termini and calculate their N-terminal acetylation yields based on the quantification values of each d<sub>0</sub> and d<sub>3</sub> species. The final “result” file is automatically created as a tabulated .csv file.

[INSERT FIGS. 2 AND 3 HERE]

#### 4 NOTES

1. The pETM41 plasmid was reconstructed from pETM41-AtNAT10 [12] by deletion of AtNAA10 (Fig. 1) and is available upon request. pETM41 is a pET derivative featuring a T<sub>7</sub> promoter under the control of the *lac* repressor, i.e., IPTG-inducible to promote overexpression of T<sub>7</sub> RNA polymerase and consequently of the MBP-NAT fusion of interest [24]. The bacterial strains used are BL21 (DE3), Rosetta (DE3) or Rosetta 2 containing plasmid pLys. Lysozyme expression from pLysS inhibits T<sub>7</sub> RNA polymerase in the absence of IPTG. This system allows stringent control, fine-tuning and high expression of the NAT gene of interest.
2. MBP (40 kDa) is expected to help refolding of the NAT of interest while not hampering or altering NAT activity due to the long in frame inter-domain linker. This linker is composed of 25 residues - including a 14 Asn-Ser stretch that is most likely unfolded - following the MBP last codon and the N-terminal Met codon within the *Nco*I restriction site of the fused NAT gene of interest.
3. *In vitro* kinetic measurements with small peptides derived from the substrates identified through the GAP assay are useful to confirm NAT specificity (see introduction and [25]). It allows also further investigation of the NAT activity towards predicted protein substrates derived from actual proteins of the species of interest. The occurrence of both the 6xHis tag and MBP allow easy purification by immobilized metal or amylose affinity resins affinity chromatography. The Tobacco Etch Virus (TEV) protease-cleavable linker of the inter-domain (Fig. 1) permits *in vitro* cleavage of MBP from the expressed NAT if the pETM41-NAT vector is used for protein purification. Upon cleavage, MBP can be removed from the NAT of interest by taking advantage of a second affinity column on the same resin. The NAT is easily recovered as it is not retained and elutes with the flow-through.

4. The GAP pipeline is useful to challenge any NAT, provided the NAT does not need an auxiliary subunit to promote or modify its activity. This is the case of both the *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* NatA complexes, which need both NAA10 and NAA15 coexpression [3]. To overproduce such complexes, pDuet vectors (Novagen) may be useful for dicistronic expression of both subunits together of the complex [26]. Nevertheless, unlike fungi NAA10, *A. thaliana* AtNAA10 was shown to display NatA full specificity without its auxiliary subunit AtNAA15 according to GAP data [19].
5. Deletion of the N-terminal signal pre-sequence - if any - of the NAT of interest is recommended. Use of predictors of sorting signals such SignalP or TargetP is useful to this end [27].
6. If necessary, the tube may be heated at 35-40 °C. Crystals may start to appear in the first hours.
7. Dry N-acetoxy-[<sup>2</sup>H<sub>3</sub>]-succinimide can be stored at -20 °C for up to a couple of years. Crystals in DMSO ready to use can also be prepared and stored at -20°C.
8. In *E. coli* the endogenous N-terminal acetylation yield decreases with growth rate and increases at the stationary phase (see Figure 5 and Supplementary Figure 11 in [15]), which explains why an analysis in rich medium and in the log phase is recommended. Such conditions favour a lower N-acetylation background.
9. The pellet can be frozen in liquid nitrogen prior to storage at -80 °C or used immediately for preparation of the protein extract.
10. Check overexpression of the MBP-NAT fusion by denaturing polyacrylamide (12%) gel electrophoresis by comparing the soluble proteins of samples before and after IPTG. An immunoblot with anti-MBP or anti-6xHis tag antibodies may be useful to confirm the expression of the NAT of interest.
11. Insert the NAT protein sequence of interest as an additional open reading frame in the *E. coli* protein database. This will be useful to check whether peptides derived from the NAT can be retrieved. This additional control allows to check expression of the NAT of interest as relevant peptides are frequently retrieved.

12. The protocol is set for volumes suitable for the treatment of 1 mg of proteins, as determined with the Bradford quantitation assay.
13. A list of the most significant 16 naturally occurring acetylated proteins out of 39 with quantitative assessment of a significant acetylation yield (>5%) is available in Table 3. These data – which may easily be obtained with the empty pETM41 or pETM41-AtNAA15 vectors analysis but – are useful to check whether the protocol is working fine. Using these two control constructs, 39 and 31 acetylated proteins were retrieved, respectively (see Supplementary Table 1 in [9] and Table 1 without quantification in [15]). A list of 23 naturally-occurring N-formylated proteins is also available in [9].
14. N-formylation can also be checked as an additional control to identify the naturally occurring N-formylated proteins. This may increase both the time and disk-storage space required for data processing and is not recommended in routine.
15. Data are from [9,15,20]

## **5 ACKNOWLEDGEMENTS**

This work was supported by KatNat (ERA-NET, ANR-17-CAPS-0001-01) and CanMore (France-Germany PRCI, ANR-20 CE92-0040) grants funded by French National Research Agency (ANR). This work has benefited from the support of a French State grant (ANR-17-EUR-0007, EUR SPS-GSR) managed by the ANR under an Investments for the Future program (ANR-11-IDEX-0003-02), from the facilities and expertise of the I2BC proteomic platform (Proteomic-Gif, SICaPS), supported by IBiSA, Ile de France Region, Plan Cancer, CNRS and Paris-Saclay University and from COST Action CA20113, supported by COST (European Cooperation in Science and Technology). We thank Lucile Jomat and Mahmoud Abdou (This team) for their contributions to the setting of robust plasmid and bacterial sample preparations, respectively.

## 6 REFERENCES

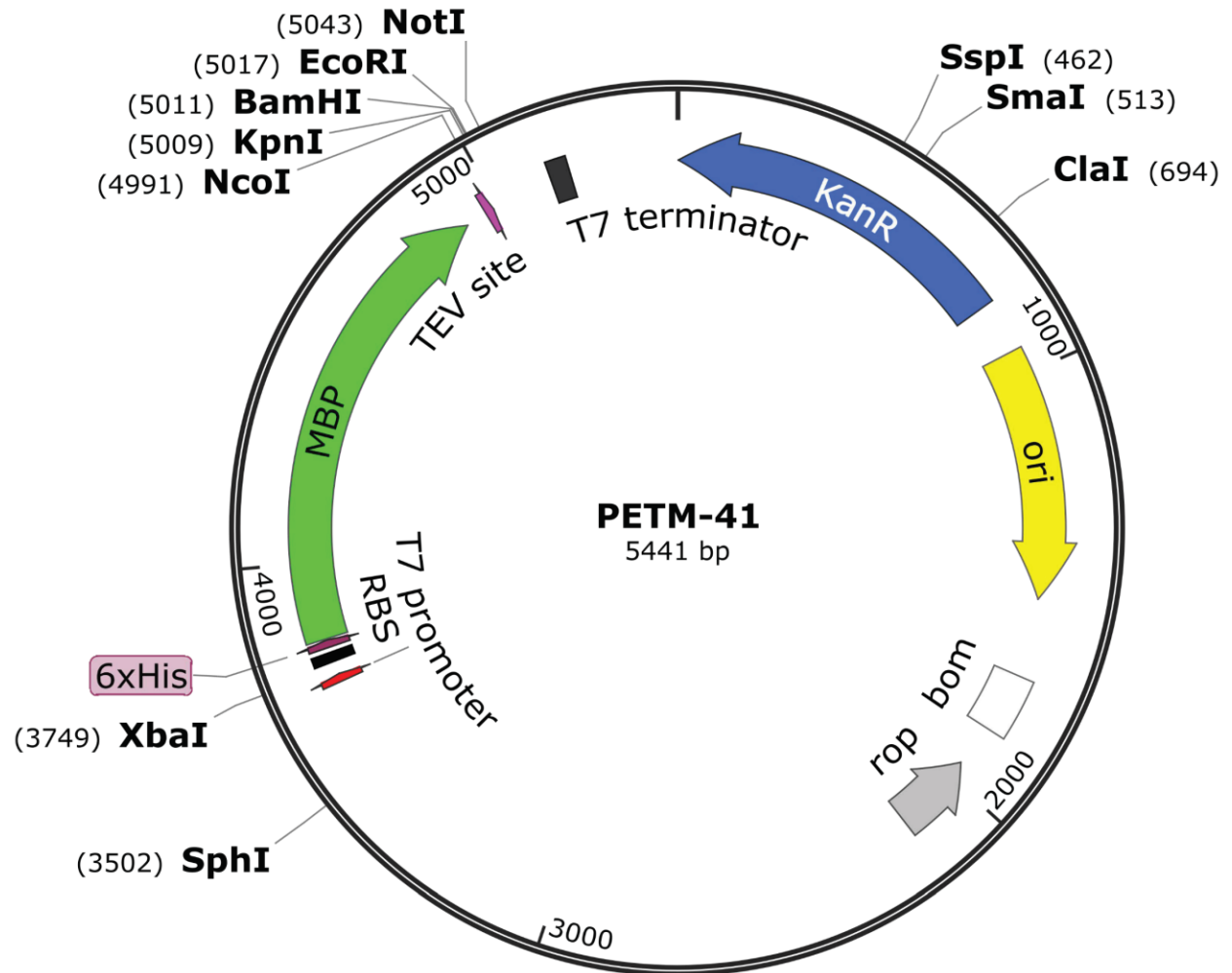
1. Bogaert A, Gevaert K (2020) Protein amino-termini and how to identify them. *Expert Rev Proteomics* 17 (7-8):581-594.
2. Starheim KK, Gevaert K, Arnesen T (2012) Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem Sci* 37 (4):152-161.
3. Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F, Gevaert K (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci U S A* 106 (20):8157-8162.
4. Van Damme P, Lasa M, Polevoda B, Gazquez C, Elosegui-Artola A, Kim DS, De Juan-Pardo E, Demeyer K, Hole K, Larrea E, Timmerman E, Prieto J, Arnesen T, Sherman F, Gevaert K, Aldabe R (2012) N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *Proc Natl Acad Sci U S A* 109 (31):12449-12454.
5. Polevoda B, Norbeck J, Takakura H, Blomberg A, Sherman F (1999) Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. *EMBO J* 18 (21):6155-6168.
6. Polevoda B, Sherman F (2001) NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. *J Biol Chem* 276 (23):20154-20159.
7. Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinnel T, Giglione C (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. *Mol Cell Proteomics* 11 (6):M111 015131.
8. Van Damme P, Evjenth R, Foyn H, Demeyer K, De Bock PJ, Lillehaug JR, Vandekerckhove J, Arnesen T, Gevaert K (2011) Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. *Mol Cell Proteomics* 10 (5):M110 004580.

9. Bienvenut WV, Giglione C, Meinnel T (2015) Proteome-wide analysis of the amino terminal status of *Escherichia coli* proteins at the steady-state and upon deformylation inhibition. *Proteomics* 15 (14):2503-2518.
10. Giglione C, Boularot A, Meinnel T (2004) Protein N-terminal methionine excision. *Cell Mol Life Sci* 61 (12):1455-1474.
11. Giglione C, Pierre M, Meinnel T (2000) Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. *Mol Microbiol* 36 (6):1197-1205.
12. Giglione C, Fieulaine S, Meinnel T (2009) Cotranslational processing mechanisms: towards a dynamic 3D model. *Trends Biochem Sci* 34 (8):417-426.
13. Giglione C, Fieulaine S, Meinnel T (2015) N-terminal protein modifications: bringing back into play the ribosome. *Biochimie* 114:134-146.
14. Dinh TV, Bienvenut WV, Linster E, Feldman-Salit A, Jung VA, Meinnel T, Hell R, Giglione C, Wirtz M (2015) Molecular identification and functional characterization of the first Nalpha-acetyltransferase in plastids by global acetylome profiling. *Proteomics* 15 (14):2426-2435.
15. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M (2016) The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* 34 (1):104-110.
16. Perez L, Ryu Y (2015) RimJ-catalyzed sequence-specific protein N-terminal acetylation in *Escherichia coli*. *Adv Bio Biotech* 6 (3):182-193.
17. Bienvenut WV, Giglione C, Meinnel T (2017) SILProNAQ: a convenient approach for proteome-wide analysis of protein N-termini and N-terminal acetylation quantitation. *Methods Mol Biol* 1574:17-34.
18. Bienvenut WV, Scarpelli JP, Dumestier J, Meinnel T, Giglione C (2017) EnCOUNTER: a parsing tool to uncover the mature N-terminus of organelle-targeted proteins in complex samples. *BMC Bioinformatics* 18 (1):182.
19. Giglione C, Meinnel T (2021) Evolution-driven versatility of N terminal acetylation in photoautotrophs. *Trends Plant Sci* 26 (4):375-391.



20. Bienvenut WV, Brünje A, Boyer J-B, Mühlenbeck JS, Bernal G, Lassowskat I, Dian C, Linster E, Dinh TV, Koskela MM, Jung V, Seidel J, Schyrba LK, Ivanauskaite A, Eirich J, Hell R, Schwarzer D, Mulo P, Wirtz M, Meinnel T, Giglione C, Finkemeier I (2020) Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. *Mol Syst Biol* 16 (7):e9464.
21. Huber M, Bienvenut WV, Linster E, Stephan I, Armbruster L, Sticht C, Layer DC, Lapouge K, Meinnel T, Sinning I, Giglione C, Hell R, Wirtz M (2020) NatB-mediated N-terminal acetylation affects growth and abiotic stress responses. *Plant Physiol* 182 (2):792-806.
22. Linster E, Layer D, Bienvenut WV, Dinh TV, Weyer FA, Leemhuis W, Brünje A, Hoffrichter M, Miklankova P, Kopp J, Lapouge K, Sindlinger J, Schwarzer D, Meinnel T, Finkemeier I, Giglione C, Hell R, Sinning I, Wirtz M (2020) The Arabidopsis N $\alpha$ -acetyltransferase NAA60 locates to the plasma membrane and is vital for the high salt stress response. *New Phytol* 228 (2):554-569.
23. Linster E, Stephan I, Bienvenut WV, Maple-Grodem J, Myklebust LM, Huber M, Reichelt M, Sticht C, Geir Moller S, Meinnel T, Arnesen T, Giglione C, Hell R, Wirtz M (2015) Downregulation of N-terminal acetylation triggers ABA-mediated drought responses in Arabidopsis. *Nat Commun* 6:7640.
24. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185:60-89.
25. Asensio T, Dian C, Boyer JB, Rivière F, Meinnel T, Giglione C (2022) A continuous assay set to screen and characterize novel protein N-acetyltransferases unveils rice General Control Non-repressible 5-Related N-Acetyltransferase 2 activity. *Front Plant Sci* 13:832144.
26. Liszczak G, Goldberg JM, Foyn H, Petersson EJ, Arnesen T, Marmorstein R (2013) Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nat Struct Mol Biol* 20 (9):1098-1105.
27. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2 (4):953-971.

## 7 FIGURES



**Figure 1: The pETM41 vector for cloning and expression of MBP-NAT fusions.**

The map was redrawn from SnapGene 5.3.2. The main features are cartooned with plain coloured arrows and the unique restriction sites and their localizations are indicated.

## Export search results

Export format

Significance threshold p <  at  identity  homology

Target FDR  
(overrides significance threshold if set)

FDR type

Ions score cut-off

Max. number of hits

Protein scoring  Standard  MudPIT

Include same-set protein hits  
(additional proteins that span the same set of peptides)

Include sub-set protein hits  
(additional proteins that span a sub-set of peptides)

Group protein families

Require bold red

Show Percolator scores

Preferred Taxonomy\*

\* Occasionally requires information to be retrieved from external utilities, which can be slow

### Protein Hit Information

Score

Description\*

Mass (Da)\*

Number of queries matched

Percent coverage\*\*

Length in residues\*\*

pI\*\*

Taxonomy\*\*

Taxonomy ID\*\*

Protein sequence\*\*

emPAI

Protein quantitation

\* Occasionally requires information to be retrieved from external utilities, which can be slow

\*\* Always requires information to be retrieved from external utilities, which can be slow

Figure 2: Mascot export parameters



## Search Information

Header	<input checked="" type="checkbox"/>
Decoy	<input checked="" type="checkbox"/>
Modification deltas	<input checked="" type="checkbox"/>
Search parameters	<input checked="" type="checkbox"/>
Format parameters	<input checked="" type="checkbox"/>
Residue masses	<input type="checkbox"/>
Quantitation header	<input checked="" type="checkbox"/>

## Peptide Match Information

Experimental Mr (Da)	<input checked="" type="checkbox"/>
Experimental charge	<input checked="" type="checkbox"/>
Calculated Mr (Da)	<input checked="" type="checkbox"/>
Mass error (Da)	<input checked="" type="checkbox"/>
Start	<input checked="" type="checkbox"/>
End	<input checked="" type="checkbox"/>
Number of missed cleavages	<input checked="" type="checkbox"/>
Score	<input checked="" type="checkbox"/>
Homology threshold	<input checked="" type="checkbox"/>
Identity threshold	<input checked="" type="checkbox"/>
Expectation value	<input checked="" type="checkbox"/>
Sequence	<input checked="" type="checkbox"/>
Frame number	<input type="checkbox"/>
Variable Modifications	<input checked="" type="checkbox"/>
Number of fragment ion matches	<input type="checkbox"/>
Query title	<input checked="" type="checkbox"/>
Peptide quantitation	<input checked="" type="checkbox"/>
Unassigned queries (peptide matches not assigned to protein hits)	<input type="checkbox"/>
Show duplicate peptides	<input checked="" type="checkbox"/>

Figure 3: Mascot search parameters



## 8 TABLES

<b>Time (min)</b>	<b>% <u>A</u></b>	<b>% <u>B</u></b>
0	100	0
5	100	0
15	74	26
40	65	35
45	40	60
46	0	100
55	0	100
56	100	0
76	100	0

**Table 1: HPLC SCX gradient**

<b>Time (min)</b>	<b>% A</b>	<b>% B</b>
0	95	5
42	65	35
43	20	80
55	20	80
55	40	60

**Table 2: nanoLC-MS/MS gradient**



Entry	Acetylated position	N-Acetylated Sequence	Average acetylation yield (%)	Gene	Function
P21170	2	SDDMSMGLPS	98	speA	Biosynthetic arginine decarboxylase
P0CE47/48	2	SKEKFERTKP	84	tufA/B	Elongation factor Tu 1/2
P0A7W1	2	AHIEKQAGEL	79	rpsE	30S ribosomal protein S5
P69805	2	VDTTQTTTEK	68	manZ	PTS system mannose-specific EIID component
P0AG86	2	SEQNNTMTF	50	secB	Protein-export protein SecB
P0A7K2	2	SITKDQIEA	46	rplL	50S ribosomal protein L7/L12
P39263	2	SAITESKPTR	29	yfcC	Putative basic amino acid antiporter YfcC
P0ABJ3	2	ATDTLTHATA	28	cyoC	Cytochrome o ubiquinol oxidase subunit 3
P0AAD8	2	STSDSIVSSQ	25	tdcC	Threonine/serine transporter TdcC
P33599	2	TDLTAQEPAW	15	nuoC	NADH-quinone oxidoreductase subunit C/D
P13009	2	SSKVEQLRAQ	14	metH	Methionine synthase
P0AEH5	2	SNQFGDTRID	8	corC	Magnesium and cobalt efflux protein CorC
P77610	2	SKHDTDTSDQ	8	ansP	L-asparagine permease
P0AAJ3	2	AMETQDIKR	6	fdnH	Formate dehydrogenase, iron-sulfur subunit
P00954	2	TKPIVFSGAQ	6	trpS	Tryptophan-tRNA ligase
P0A9C5	2	SAEHVLTMLN	5	glnA	Glutamine synthetase

**Table 3:** Major N-terminal acetylated proteins from *E. coli*.