Dissertation

Submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

presented by
Msc. Trinh Dinh-Van
born in Nghe An province, Vietnam

Oral-examination: December 16th, 2013

Missing links of the protein N^{α}-terminal acetylation machinery in plants

Referees: Prof. Dr. Rüdiger Hell

Prof. Dr. Sabine Strahl

“What we know is a drop, what we don't know is an ocean”

Sir Isaac Newton

Table of contents

General statement VI
Acknowledgement VII
List of abbreviation VIII
Summary X
1 INTRODUCTION 1
1.1Protein Nt-acetylation 1
1.1.1 N-terminal acetylation and N-terminal acetyltransferases 1
1.1.2 Functions of N-terminal acetylation 2
1.2 N-terminal acetylation in prokaryote 3
1.2.1 N-terminal acetylation in bacteria 3
1.2.2 N-terminal acetylation in Archaea 4
1.3 N-terminal acetylation in yeast 5
1.3.1 Yeast NatA 5
1.3.2 Yeast NatB 6
1.3.3 Yeast NatC 7
1.3.4 Yeast NatD 7
1.3.5 Yeast NatE 8
1.4 N-terminal acetylation in human 8
1.4.1 Human NatA 9
1.4.2 Human NatB 10
1.4.3 Human NatC 10
1.4.4 Human NatD 11
1.4.5 Human NatE 12
1.4.6 Human NatF 12
1.5 N-terminal acetylation in plants 13
1.5.1 Identified NATs complexes 13
1.5.2 Evidences of protein N^{α}-acetylation in plant organelles 15
1.6 Aims of project 17
2. MATERIAL AND METHODS 18
2.1 Technical equipment and materials 18
2.1.1 Technical equipment 18
2.1.2 Chemicals 19
2.1.3 Consumables 21

2.1.4 Kits 21
2.1.5 Enzymes and antibodies 21
2.1.6 Primers 22
2.1.7 Softwares 25
2.1.8 Buffers and solutions 26
2.1.9 Oligo peptides 27
2.2 Plant materials and methods 27
2.2.1 Plant materials 27
2.2.2 Growth conditions 28
2.2.3 *In vitro* growth of Arabidopsis seedlings on agar plates 28
2.2.4 Drought stress treatment 29
2.2.5 UV-c tress treatment 29
2.2.6 Arabidopsis protoplast isolation 29
2.2.7 PEG-Calcium transfection of Arabidopsis protoplasts 30
2.2.8 Osmotic plasma membrane lysis of protoplasts 30
2.2.9 Infiltration of tobacco leaves 30
2.3 Bacteriological Methods 30
2.3.1 Bacterial strains 30
2.3.2 Bacterial growth 31
2.3.3 Preparation of electro-competent cells 31
2.3.4 Preparation of Agrobacterium competent cells 31
2.3.5 Electro-transformation of bacteria 31
2.3.6 Heat-shock transformation of Agrobacterium 32
2.3.7 Plasmids isolation of bacteria 32
2.3.8 Bacterial glycerol stock 32
2.3.9 Total protein extraction from *E. coli* *32*
2.4 Molecular Biology Methods 32
2.4.1 Genomic DNA extraction from *Arabidopsis thaliana* 32
2.4.2 Polymerase chain reaction (PCR) 33
2.4.3 DNA gel electrophoresis 34
2.4.4 Cloning using Endonucleases and Ligase 34
2.4.5 Gateway cloning 35
2.4.6 Total RNA extraction from *Arabidopsis thaliana* 35
2.4.7 cDNA synthesis 35

2.4.8 qRT-PCR35
2.5 Protein methods 36
2.5.1 Total protein extraction from *Arabidopsis thaliana* 36
2.5.2 Purification of total protein extract 36
2.5.3 His-tag protein purification 36
2.5.4 *In vitro* Acetyltransferase assay 36
2.5.5 *In vitro* auto lysine acetylation assay 37
2.5.6 NBD-Cl fluorescence assay 37
2.5.7 Preparation and analysis of membrane fractions 37
2.5.8 Protein SDS-PAGE, Western blot and immunodetection 38
2.5.9 LC-MS/MS analysis 38
2.6 Microscopic methods 39
2.6.1 Counting of protoplasts 39
2.6.2 CLSM Analysis 39
3. RESULTS 40
3.1 Identification and characterization of Arabidopsis NatD 40
3.1.1 Identification of Arabidopsis N^{α}-acetyltransferase Naa40p by sequence similarity . 40
3.1.2 Over-expression and purification of recombinant AtNaa40p 42
3.1.3 Arabidopsis Naa40p displays N^{α}-acetyltransferase activity 43
3.1.4 Subcellular localization of AtNaa40p 45
3.1.5 Identification and characterization of atnaa40 mutants 48
3.1.5.1 Identification of *atnaa40* mutants 48
3.1.5.2 The transcription level of ATNAA40 in *atnaa40-1* and *atnaa40-2* are altered 49
3.1.5.3 Phenotypes of *atnaa40-1* and *naa40-2* mutants 50
3.1.5.4 Phenotypes of *atnaa40* mutants under UV-c stress 52
3.1.5.6 Status of protein N-terminal acetylation in the *atnaa40* mutants 54
3.2 Identification and characterization of Arabidopsis NatF 55
3.2.1 Identification of Arabidopsis N^{α}-acetyltransferase Naa60p by sequence similarity . 55
3.2.2 Over-expression and purification of recombinant AtNaa60p 57
3.2.3 Arabidopsis Naa60p displays N^{α}-acetyltransferase activity 58
3.2.4 Subcellular localization of AtNaa60p 59
3.2.5 AtNaa60p potentially associates with another membrane-anchored protein 65
3.2.6 Identification and characterization of *atnaa60* mutant 67
3.2.6.1 Identification of *atnaa60* mutant 67

3.2.6.2 The expression of ATNAA60 is severely disrupted in *atnaa60* mutant 68
3.2.6.3 Phenotypes of *atnaa60* mutant 67
2.3.6.4 Germination of *atnaa60* seeds in different media 69
2.3.6.5 Phenotypes of *atnaa60* seedlings under stress conditions 71
2.3.6.6 Status of protein N-terminal acetylation in the *atnaa60* mutant 72
3.3 Identification of plastidic N^{α}-acetyltransferases (pNATs) 74
3.3.1 Searching for the putative pNATs 74
3.3.2 Subcellular localization of putative pNATs 74
3.3.3 *In silico* analysis of putative pNATs 76
3.3.4 Overexpression of putative pNATs in *E. coli* 78
3.3.5 LC-MS/MS analysis of Nt-acetylation by pNATs 80
3.3.5.1 Nt-acetylation by At2g39000p (pNaa10p) 80
a. Identification of recombinant At2g39000p in the sample 80
b. Substrate specificity of At2g39000p (pNaa10p) 81
c. pNaa10p acetylates lysine residues of itself 83
3.3.5.2 Nt-acetylation by At1g24040 (pNaa20p) 84
a. Identification of recombinant At1g24040p in the sample 84
b. Substrate specificity of At1g24040p (pNaa20p) 84
3.3.5.3 Nt-acetylation by At2g06025 (pNaa30p) 86
a. Identification of recombinant At2g06025p in the sample 86
b. Substrate specificity of At2g06025p (pNaa30p) 87
3.3.6 Identification and characterization of T-DNA insertion mutants of pNATs 88
3.3.7 Expression of plastidic N^{α}-acetyltransferase coding genes under drought condition 90
4. DISCUSSION 92
4.1 Protein Nt-acetylation by AtNatD 92
4.1.1 AtNaa40p localizes to the cytoplasm and to the nucleus 92
4.1.2 N^{α}-acetyltransferase activity of NatD 93
4.1.3 Histone N-tail and phenotypes of *naa40* mutant 95
4.1.4 Substrates of NatD 96
4.2 Protein Nt-acetylation by AtNatF 97
4.2.1 Is the auxiliary subunit of AtNatF a membrane-anchored protein? 97
4.2.2 N^{α}-acetyltransferase activity of AtNaa60p 97
4.2.3 Relation between localization and phenotypes of atnaa60 mutant 98
4.2.4 NatF N-terminally acetylates numerous proteins 99

4.3 Protein Nt-acetylation by AtpNATs 99
4.3.1 Bacterial proteins – an unrestricted source for Nt-acetylation 99
4.3.2 Nt-acetylation in chloroplast and plastidic NATs..... 101
4.3.3 N^{α}-acetyltransferase activity of pNATs..... 102
4.3.4 Acetylation of lysine residues is necessary for efficient function of pNaa10p 103
4.3.5 Are more than three NATs responsible for acetylation in chloroplast? 104
4.3.6 Role of chloroplastic protein acetylation in drought stress response..... 105
4.4 Nt-acetylation in *Arabidopsis thaliana*, are there more NATs? 106
4.5 The overall picture of protein N-terminal acetylation in plants..... 108
5. REFERENCES..... 109
Subpplemental data 123

GENERAL STATEMENT

I declare that I am the sole author of this submitted dissertation and that I did not make use of any sources or help apart from those specifically referred to. Experimental data or material collected from or produced by other persons is made easily identifiable. I also declare that I did not apply for permission to enter the examination procedure at another institution and that the dissertation is neither presented to any other faculty, nor used in its current or any other form in another examination.

_____ _____

(Place, Date) (Trinh Dinh-Van)

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Rüdiger Hell for the support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervision and mentor for my Ph.D study.

Besides my supervisor, I would like to thank the rest of my TAC committee: Prof. Dr. Sabine Strahl, Prof. Dr. Matthias Mayer for their ideas, encouragement, insightful comments, and hard questions.

My sincere thanks go to Dr. Markus Wirtz for his inspiring guidance, generous assistance, constructive critical suggestions and consistent advice throughout this study. I am also very thankful to our collaboration partner: Dr. *Willy Bienvenut, Dr.* Carmela Giglione, Dr. Thierry Meinnel (ISV-Gif, NCRS, France), without you the chloroplastic NATs have never been found.

I am greatly thankful to Ministry of Education and Training of Vietnam (MOET) for providing financial support for the project.

I thank my fellow labmates in AG Hell: Anna Speiser, Hannah Birke, Monika Huber, Ilaria Forieri, Carolin Seyfferth, Marleen Silbermann, Eric Linster, Alex Lee, Arman Allboje, Florian Haas, Stefan Haberland, Gernot Poschet, Michaels3 (Schulz, Kraft and Schilbach), Marcel Beier, Stefan Greiner, Birgit Maresch, Angelika Wunderlich and Inge.

Last but the most, it has never been enough to mention my feelings, my deep gratitude for my parents who give me life, teach me how to live, how to love and how to become a good person. My sister, my brother and I, we have never said it out loud but we all know how much we love each other and how much we mean to each other. Thanks god for giving me the best family in the whole world.

LIST OF ABBRIVIATIONS

Ac-CoA	acetyl coenzymeA
Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
At	*Arabidopsis thaliana*
BSA	bovine serum albumin
cDNA	complementary DNA
CLSM	confocal laser scanning microscopy
Col-0	Arabidopsis thaliana ecotype Columbia-0
Cpm	counts per minute
emPAI	exponentially modified protein abundance index
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide solution mix
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EGTA	glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
ER	endoplasmic reticulum
EYFP	enhanced YFP
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
GNAT	GCN5-related N-acetyltransferase
h	human
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
Ile (I)	isoleucine
IPTG	isopropyl-D-1-thiogalactopyranoside
LC-MS/MS	liquid chromatography–mass spectrometry
Leu (L)	leucine
Lys (K)	lysine
MAP	methionine aminopepridase

MBP	maltose binding protein
Met (M)	methionine
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Murashige & Skoog
NAT	N^{α}-acetyltransferase
NBD-Cl	4-Chlor-7-nitrobenzo-2-oxa-1,3-diazol
NLS	nuclear localization signal
Nt-acetylation	N-terminal acetylation
PAGE	polyacrylamide gelelectrophoresis
PCR	polymerase chain reaction
PEG	polyethylenglycol
Phe (F)	phenylalanine
PM	plasma membrane
PMSF	phenylmethanlsulphonylfluoride
Pro (P)	proline
qRT-PCR	quantitative real time polymerase chain reaction
RFP	red fluorescent protein
RNA	ribonucleic acid
RT	room temperature
SDS	sodiumdedocylsulfate
Ser (S)	serine
TAIR	the Arabidopsis information resource
TBS	tris buffered saline
TCA	trichloroacetic acid
T-DNA	transferred DNA used for insertional mutagenesis
TEMED	N,N,N',N'-Tetramethylethylenediamine
Thr (T)	threonine
TMD	transmembrane domain
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
Tyr (Y)	tyrosine
UV	ultra violet
Val (V)	valine
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
y	yeast
YFP	yellow fluorescent protein

SUMMARY

Protein N-terminal acetylation (Nt-acetylation) is the transfer of acetyl group from acetyl coenzyme A (Ac-CoA) to the alpha amino acid of a protein. Since it has been discovered more than fifty years ago, Nt-acetylation is known to be one of the most common protein modifications in eukaryotes, occurring on approximately 50-70% of yeast soluble protein and about 80-90% of human protein. However, the exact biological role has remained enigmatic for majority of affected proteins, and only for a small number of proteins, Nt-acetyation was linked to various features of protein such as localization, stability and interaction. Nt-acetylation in yeast and in human is thoroughly investigated with the identification of five (NatA-NatE) and six (NatA-NatF) N^{α}-acetyltransferase (NAT) types, respectively. In contrast, the knowledge of Nt-acetylation in plants was vacant for many years. The first Arabidopsis NAT, AtNatC was identified in 2003, and very recently three more NATs (NatA, NatB and NatE) were described by Iwona Stephan. In this study, we identified two NATs (NatD and NatF) that are still missing in plants. AtNatD/AtNaa40p is conserved from yeast with respect to acetylation of protein histone H4. The lack of N-terminal serine acetylation increases the overall positive charge of H4 N-tail which causes the minor phenotypes observed in *atnaa40* mutant. The acetylation of N-terminal serine of histone H4 might also involve in DNA double-strand break response. Besides, the subcellular localization to cytoplasm and nucleus suggests a lysine acetyltransferase activity of AtNaa40p towards histones. AtNatF/AtNaa60 unusually localizes to plasma membrane and to the tonoplast. The sensitivity of *atnaa60* mutant to salt tress during germination stage appears to be related to the localization, and indicating the involvement of AtNaa60p in salt stress or osmotic stress response. Like hNaa60p, AtNaa60 is believed to acetylate a large number of proteins according to the NBD-Cl fluorescent assay. AtNaa60p acetylates methione and serine starting peptides *in vitro*.

In addition, numerous proteins are found N-terminally acetylated in chloroplasts, both chloroplast-encoded and nuclear-encoded proteins. *In silico* study reveals eight putative plastidic NATs of which seven localize to the chloroplasts when they are transiently expressed with EYFP in Arabidopsis protoplasts. Three proteins (At2g39000, At1g24040 and At2g06025) acetylate plenty of *Escherichia coli* proteins, their substrate specificities are strongly correlated to chlotoplast transit peptide (cTP) cleavage sites. Four other proteins (At4g19984, At1g26220, At1g32070 and At4g28030) are possibly true NATs since they possess the conserved Ac-CoA binding motif. Our results, together with other studies on acetylation in chloroplast, propose the connection between Nt-acetylation of chloroplastic proteins and drought stress.

ZUSAMMENFASSUNG

Die N-terminale Acetylierung von Proteinen, ist der Transfer einer Acetyl-Gruppe von Acetyl Coenzym A auf die alpha Aminogruppe eines neu synthetisierten Proteins. N-terminale Acetylierung ist eine der am häufigsten vorkommenden Proteinmodifikationen in Eukaryoten. In Hefe sind 50 – 70 % und im Menschen sogar 80 – 90 %aller löslichen Proteinen am N-alpha Terminus acetyliert. Trotz der hohen Acetylierungsrate ist die genaue biologische Funktion der N-terminalen Acetylierung weit gehend unbekannt. Nur für eine geringe Anzahl von Proteinen ist die Funktion der Acetylierung bekannt, wo bei die Modifikation für die Proteinlokalisierung, -stabilität und -interaktion benötigt wird. Am besten erforscht ist die N-terminale Acetylierung in Hefe und Mensch. In Hefe sind fünf (NatA – NatE) und im Menschen sechs (NatA – NatF) N^{α}-acetyltransferase (NAT) Komplexe identifiziert und charakterisiert. Der erste Arabidopsis NAT Komplex, AtNatC, wurde erst im Jahre 2003 beschrieben. Drei weitere NATs (NatA, NatB und NatC) wurden von Iwona Stephan charakterisiert.

In dieser Arbeit wurden zwei weitere Arabidopsis NATs (NatD und NatF) identifiziert. Die Acetylierungsaktivität von AtNatD/AtNaa40p ist konserviert von Hefe bis Arabidopsis, so werden die Histonproteine H4 durch AtNaa40p acetyliert. Der Verlust der N-terminalen Actylierung von Serinen führt zu einer Erhöhung der positiven Ladung am N-terminus von Histon H4. Diese veränderte Ladung von Histon H4 hat den Wachstums-Phänotyp von *atnaa40* zur Folge. Die fehlende Acetylierung von Histon H4 ist vermutlich auch in die DNA-Qualitätskontrolle, in der Antwort auf einen DNA-Doppelstrangbruch, involviert. Neben dem Zytoplasma wurde AtNaa40p auch im Nucleus lokalisiert. Die Lokalisierung von AtNaa40 im Nukleus könnte auf eine Lysine Acetyltransferase Aktivität an Histon H4 hindeuten. Unerwarteter Weise wurde AtNatF/AtNaa60p in der Plasmamembran und im Tonoplasten detektiert. Diese Lokalisierung könnte mit der Sensitivität der *atnaa40* Mutante gegen Salz während der Keimung zusammenhängen und deutet auf eine Verbindung von AtNaa60 und der zellulären Antwort auf Salzstress oder osmotischen Stress hin. Wie hNaa60 ist wahrscheinlich auch AtNaa60p für die Acetylierungeiner großen Anzahl von Proteinen verantwortlich. Dies wurde durch die Quantifizierung mit NBD-Cl gezeigt. AtNaa60 acetyliert *in vitro* Peptide die mit einem Methionin oder Serin beginnen.

Zusätzlich wurden zahlreiche plastidären Proteine entdeckt, die N-terminalacetyliert waren. Bei diesen Proteinen handelt es sich sowohl um plastidär als auch kerncodierte Proteine. Mit *in silico* Analysen wurden acht putativ plastidäre NATs identifiziert, von denen die Lokalisierung im Chloroplasten mit Hilfe von transient exprimierten EYFP in

Arabidopsis Protoplasten bestätigtwurde. Drei Proteine (At2g39000, At1g24040 und At2g06025) acetyliereneine große Anzahl von *Escherichia coli* Proteinen, deren Substratspezifität stark mit der cTP Schnittstelle korreliert. Vier weitere Proteine (At4g19984, At1g26220, At1g32070 und At4g28030) sind wahrscheinlich vollwertige NATs, da sie alle die konservierte Ac-CoA Bindungsdomäne besitzen. Unsere Ergebnisse postulieren, zusammen mit anderen Studien an N-terminaler Acetylierung in Chloroplasten, eine Verbindung von N-terminalen Acetylierung plastidärer Proteine mit Trockenstress.

1. INTRODUCTION

1.1 Protein N-terminal acetylation

1.1.1 N-terminal acetylation and N-terminal acetyltransferases

N-terminal acetylation (Nt-acetylation) of protein is an enzymatic process which transfers the acetyl group from acetyl coenzyme A (Ac-CoA) to the alpha amino acid residue of a protein. The acetylated amino acid is the initial methionine or the second residue after cleaving off methionine by methionine aminopeptidase (MAP). Among the other modifications of N-terminus of protein such as formylation, methylation, propionylation, glycosylation and others Nt-acetylation is the most common modification (Driessen et al., 1985). More than 80% of cytosolic proteins in human and approximately 90% of the protein from mouse L cells are N^{α}-acetylated (Jornvall, 1977; Jornvall et al., 1980). In lower eukaryotes (*Saccharomyces fraglis, Neurospora crassa*), about 50% of cytosolic proteins are acetylated (Brummel et al., 1971). In contrast to eukaryotes, in bacteria only few proteins were known to be acetylated (Polevoda and Sherman, 2003). However, in the archaea, Nt-acetylation seems to have widespread affecting while about 15% of proteins are acetylated (Falb et al., 2006).

Up to now Nt-acetylation is found to be irreversible. It is catalyzed by an enzyme called N^{α}-acetyltransferase (NAT), which is a member of GNAT family protein (GCN5-related N-acetyltransferase). The common N^{α}-acetyltransferase is a complex that is composed of a catalytic subunit and an auxiliary subunit. However, N^{α}-acetyltransferases can contain additional auxiliary subunit. Furthermore, the catalytic subunit can act without auxiliary subunit (Polevoda et al., 2009a). Although the first N-terminal acetylated polypeptide was discovered in 1958 (Auffret and Williams, 1978), only many years later the first N^{α}-acetyltransferase was identified in yeast (NatA, a major NAT composed of the catalytic subunit Naa10p and the auxiliary Naa15p) by Sherman and coworkers (Mullen et al., 1989). After the identification of NatA, four other NATs (NatB, NatC, NatD and NatE) were identified one after the other and shown to have different substrate specificity (Tercero and Wickner, 1992; Polevoda et al., 1999; Polevoda and Sherman, 2001; Polevoda et al., 2003; Singer and Shaw, 2003; Song et al., 2003; Gautschi et al., 2003).

In human, all orthologous NATs from yeast were found and demonstrated to be conserved regarding subunit composition, subcellular localization and substrate specificity (Ametzazurra et al., 2009; Arnesen et al., 2005; Arnesen et al., 2006; Arnesen et al., 2009;

Hole et al., 2011). During the evolution, higher eukaryotic organisms additionally express another type of NAT, NatF. Human NatF shows very distinct substrate specificity in comparison with other NATs and it contributes about 10% of overall protein Nt-acetylation in the cell (Van Damme et al., 2011c).

In eukaryotic organisms, Nt-acetylation of protein very common occurs cotranslationally when there are between 20-50 residues protruding from the ribosome (Strous et al., 1974; Strous et al., 1973). Besides, a number of proteins were found post-translationally acetylated in yeast, in human (Helsens et al., 2011; Van Damme et al., 2011b) and in plant chloroplast (Kleffmann et al., 2007; Zybailov et al., 2008; Bienvenut et al., 2012). In contrast to eukaryote, Nt-acetylation occurs post-translationally in bacteria (Gordiyenko et al., 2008).

1.1.2 Functions of N-terminal acetylation

Although thousands of proteins are N-terminal acetylated, the general function of Nt-acetylation has not been assessed. Only for a small number of acetylated proteins Nt-acetyation was linked to various features of protein. Furthermore, the knock down mutants of certain NAT can lead to lethality, abnormal growth and development or having no effect on the phenotype (Gautschi et al., 2003; Polevoda and Sherman, 2003; Polevoda et al., 2009b; Stephan, 2011).

Nt-acetylation affects protein stability

For many years, it has been thought that the general function of Nt-acetylation is to protect proteins from degradation (Jornvall, 1975; Persson et al., 1985). Experimental data indicated that proteins with acetylated N-termini are more stable than non-acetylated proteins (Hershko et al., 1984). It was explained that some proteins can become N-terminally ubiquitinated and subsequently degraded. Nt-acetylation prevents ubiquitination so that increases protein stability (Kuo et al., 2004). However, recently Varshavsky has shown that Nt-acetylation unstablize proteins (Varshavsky, 2008; Hwang et al., 2010). He found that N-terminal acetylated amino acid sequences in certain proteins can create degradation signals that is recognized by an ubiquitin ligase Doa10 and marks them with ubiquitin for degradation.

Nt-acetylation influences localization of protein

Besides having an impact on protein stability, evidences suggest that Nt-acetylation can direct proteins to their destinations. Toc159, a receptor protein that localizes at chloroplast membrane, is responsible for the transport of proteins into chloroplast. Lack of Toc159 resulted in accumulation of unprocessed N-terminal acetylated precursor proteins outside of plastids (Bischof et al., 2011). These data indicate that N-terminal acetylated termini can be path of the signal for importing proteins into chloroplast via Toc195 pathway. Not only influencing on the import of proteins into choloroplast, Nt-acetylation can also change the target of proteins in the secretory pathway. Proteins that are destined for the secretory pathway as usual contain an N-terminal signal sequence which targets them to the endoplastic reticulum (ER). N-terminal acetylated termini therefore interrupt this pathway. Forte and coworkers demonstrated that Nt-acetylation prevents translocation to the ER, however how exactly Nt-acetylation blocks protein translocation is not fully understood (Forte et al., 2011).

Apoptosis links to Nt-acetylation

Recent discoveries have supposed the links between apoptosis and Nt-acetylation. Naa10p is a catalytic subunit of major N^{α}- acetyltransferase NatA. One point mutant (p.Ser37Pro) of human Naa10p resulting in the reduction of acetylation level by NatA, caused the lethal X-linked disorder of infancy (Rope et al., 2011). In support of this theory, Arnesen and coworkers showed that the depletion of human NatA complex induces p53-dependent apoptosis and p53-independent growth inhibition (Gromyko et al., 2010). In contrast, Yi et al. (2011) demonstrated that Nt-acetylation induces apoptosis. They also proposed that Ac-CoA might serve as a signaling molecule that couples apoptotic sensitivity to metabolism by regulating protein N^{α} -acetylation.

Besides effecting protein stability, protein localization and apoptosis, recent data have shown the involvement of Nt-acetylation in other features of protein such as protein synthesis, protein-protein interaction and gene silencing (Kamita et al., 2011; Yi et al., 2011; Van Damme et al., 2011a; Arnesen 2011). However, the exact impact of Nt-acetylation on these processes is not yet understood.

1.2 N-terminal acetylation in prokaryote

1.2.1 N-terminal acetylation in bacteria

Acetylation of alpha amino group at the N-termini of proteins is very common in eukaryotes. In contrast, Nt-acetylation of bacterial proteins is rare, only seven proteins are known to be acetylated in *Escherichia coli* (*E. coli*) (Vetting et al., 2008). However, three *E. coli* NATs RimL, RimJ and RimI have been identified, they are responsible for the acetylation of three ribosomal proteins L12, S5 and S18 (Yoshikawa et al., 1987; Tanaka et al., 1989). The acetyltransferases responsible for Nt-acetylation of other proteins are currently unknown.

Unlike most of eukaryotic NATs that consist of two or more subunits, bacterial NATs appear to have no auxiliary subunit and act post-translationally on their target substrates (Vetting et al., 2005; Miao et al., 2007; Vetting et al., 2008). It has been thought that bacterial NATs have only one substrate, Vetting even gave a model of RimJ to explain the unique substrate specificity of NATs (Vetting et al., 2005). However, recently Fang et al. (2009) showed that RimJ can N-terminally acetylate both ribosomal protein L12 and Thymosin α1 (Tα1).

Regardless of the fact that the Nt-acetylation in bacteria is rare, it seems to be important for the functions of ribosome. In seven proteins that are known be acetylated, three of them are ribosomal proteins. Furthermore, acetylation of L12 increases the interaction of L7 (a N^{α}-acetylated derivative of L12) and L10 in the *E. coli* ribosomal stalk complex (Gordiyenko et al., 2008).

1.2.2 N-terminal acetylation in archaea

The first discoveries of protein Nt-acetylation in *haloarchaea* was 20 years ago, a few ribosomal proteins were reported to be acetylated at their N-termini (Kimura et al., 1989; Hatakeyama, 1990; Klussmann et al., 1993). Since then, it was thought that Nt-acetylation of archaeal proteins is similar to bacteria and therefore there was no increasing of interest in protein acetylation in archaea.

However, recently a large-scale proteomic studies on *Halobacterium salinarum* and *Natronomonas pharaonis* surprisingly revealed that Nt-acetylation is not uncommon in haloarcheae affecting 14% to 19% of soluble proteins (Aivaliotis et al., 2007; Falb et al., 2006). The acetylation occurs after methionine removal and only serine and alanine residues are acetylated. This pattern of acetylation would be consistent with NatA complex but not bacterial NATs.

The further study on *Sulfolobus solfataricus* found one gene coding a putative NAT (sso0209), which named SsArd1, shared 37% identity with human NatA catalytic subunit Ard1 (Mackay et al., 2007). Surprisingly, the characterization of various mutants revealed the unique substrate specificity of SsArd1. Besides displaying NatA-like activity (Ala and Ser residues are acetylated), SsArd1 also acetylates NatB and NatC substrates (Met-Glu- and Met-Leu- N-termini). For these reasons, it was postulated that the situation of *Sulfolobus* represents an ancestral state with a single NAT, which is not part of an enzyme complex and has broader substrate specificity than eukaryotic NATs. Eukaryotic NATs have later experienced gene duplications and have evolved further into specialized proteins. Non-catalytic subunits that may influence protein specificity and anchor the eukaryotic NAT proteins to the ribosome have also evolved (Mackay et al., 2007).

1.3 N-terminal acetylation in yeast

In the lower eukaryotic yeast, about 50% of cytosolic proteins are N-terminally acetylated. So far, five NAT complexes have been identified and denoted NatA-NatE. They differ from the others in subunit composition and substrate specificity.

1.3.1 Yeast NatA

Yeast NatA complex is one of the three major NATs, consists of catalytic subunit yNaa10p (yArd1p) and auxiliary subunit yNaa15p (yNat1p). The catalytic subunit yArd1p was first described to have important functions in cell growth, cell cycle and the mating process (Whiteway and Szostak, 1985; Whiteway et al., 1987). Later, *yARD1* coding product was shown to have Nt-acetyltransferase activity (Lee et al., 1989; Park and Szostak, 1990). Accidently, during a screening for a yeast mutant defective in histone acetyltransferase activity, Mullen and coworkers found the auxiliary yNatA subunit yNat1p (Mullen et al., 1989).

yARD1 and *yNAT1* are distinct genes that encode different proteins but are both required for N^{α}-acetyltransferase acitivity. Concurrent over-expression of both *yARD1* and *yNAT1* in yeast caused a 20 fold increase in acetyltransferase activity *in vitro*, whereas over-expression of either *yNAT1* or *yARD1* alone did not raise activity over the basal level (Mullen et al., 1989; Park and Szostak, 1992). Besides, the *ynat1* mutant exhibited the same phenotypes as *yard1* mutant (Mullen at al., 1989). Taken together, it was suggested

that the *yNAT1* and *yARD1* gene products are subunits of an acetyltransferase in yeast (Mullen et al., 1989). It was indeed, Park and Szostak (1992) showed that yArd1p functions come after *yNAT1* transcription and through its C-terminal region yArd1p forms a complex with yNat1p. Later, it was reported that the yeast N^{α}-acetyltransferase NatA is anchored to the ribosome via yNat1p and interacts with nascent polypeptides (Gautschi et al., 2003). Therefore, the acetylation of proteins by yNatA occurs co-translationally in the cytosol.

NatA is a major NAT in yeast and is expected to acetylate over 2000 proteins, of those about 100 substrates have been identified. yNatA predominantly acetylates Ser-, Thr- and Ala- N-terminal residues. Acetylation is rare for Cys-, Gly- and Pro- residues. Met- and Val- residues are not acetylated (Polevoda et al., 1999; Polevoda and Sherman, 2003; Polevoda et al., 2009a; Perrot et al., 2008).

Despite the fact that yNatA is major complex and it acetylates broad range of proteins, the yeast *ard1* or *nat1* knockout mutants are not lethal. Both *ard1* and *nat1* mutants showed slow growth, inability to sporulate and failure to enter G_o but are still viable. The double mutant did not exhibit additional phenotypes (Mullen et al., 1989).

1.3.2 Yeast NatB

Yeast NatB complex contains the catalytic subunit yNaa20p (yNat3p) (Polevoda et al. 1999) and auxiliary subunit yNaa25p (yMdm20p) (Singer et al., 2000; Polevoda et al., 2003). Via auxiliary subunit yMdm20p, NatB associates with ribosome and co-translationally transfers acetyl group to the N-termini of target proteins. Along with NatA, NatB is a major NAT and it is responsible for about 15% of all protein Nt-acetylation in yeast (Van Damme et al., 2012). It N-terminally acetylates Met followed by Glu-, Asp-, Asn-, Gln- and Met- (Polevoda et al., 2003; Van Damme et al., 2012).

The yeast *naa20-Δ* and *naa25-Δ* deletion mutants display the most prominent phenotypes in comparison to all NATs deletions, including temperature sensitivity, increase of sensitivity to salt, osmotic agents, anti-microtubele and other chemicals (Polevoda and Sherman, 2003). In further studies, *naa20-Δ* and *naa25-Δ* deletion strains showed reduced mating efficiency, abnormal morphology, random polarity in budding, defects in mitochondria division and vacuolar segregation (Polevoda et al., 2003). Interestingly, these phenotypes are similar to phenotypes of certain *act1* and *tpm1* mutant strains. Act1 and

Tpm1 are both substrates of NatB and the interaction between Actin-Tropomyosin is required for many aspects of cell division in yeast (Singer et al., 2000). Therefore, it was suggested that the phenotypes of *naa20-Δ* and *naa25-Δ* mutants are the results of the reduction in acetylation of actin and tropomyosins (Polevoda et al., 2003). In support of this theory, recently experimental data showed that yMdm20p functions with yNat3p to acetylate Tpm1 protein and regulate tropomyosin-actin interactions (Singer and Shaw, 2003; Van Damme et al. 2012).

1.3.3 Yeast NatC

Yeast NatC complex composed of the catalytic subunit yNaa30p (yMak3p) and two auxiliary subunits yNaa35p (yMak10p) and yNaa38p (yMak31p). The catalytic subunit yMak3p, which has a significant homology to the *E. coli* N^{α}-acetyltransferase RimJ, was first identified by Tercero and Wickner (1992). Then Polevoda and Sherman (2001) further characterized yNatC by identification and characterization of the additional subunits yMak10p and yMak31p. All three subunits are mandatory for the activity of yNatC. Yeast NatC is likes other yeast NATs, it associates with ribosome and co-translationally acetylates proteins in cytoplasm (Polevoda et al., 2008).

In contrast to yNatA and yNatB, yNatC substrates are quite rare. yNatC was reported to acetylate Met followed by a hydrophobic aminoacid residue (Leu-, Ile-, Phe-, Trp- (Polevoda et al., 1999; Polevoda and Sherman, 2003). Although having limited substrates, genetic depletion of any one of three subunits resulted in abnormal phenotypes. The yeast mutant strains *mak3*-Δ, *mak10*-Δ, and *mak31*-Δ displayed the similar phenotypes including the lack of acetylation of yNatC substrates *in vivo*, diminished growth at 37^{o}C on the media containing non-fermentable carbon sources, defective L-A virus propagation, and loss of telomere elongation (Polevoda and Sherman, 2003; Askree et al., 2004).

1.3.4 Yeast NatD

yNatD, the last NAT was identified in yeast, is the new type of NAT displaying several differences in comparison to the others (Song et al., 2003; Polevoda et al., 2009b). yNatD has only catalytic subunit yNaa40p (yNat4p) and does not require interaction with any protein for activity. yNatD solely expressed in a bacterial system was sufficient to acetylate its substrates (Song et al., 2003; Polevoda et al., 2009b). However, similar to the other

NATs, yNatD is associated with ribosomes and the acetylation by yNatD occurs co-translationally in cytosol (Polevoda et al., 2009b).

yNatD was reported to be responsible for the acetylation of the serine residues at the N-termini of histones H4 and H2A (Song et al., 2003). Very usual, the proteins with serine residue at the N-terminus are acetylated by NatA but not histones H4 and H2A. In yeast *ard1* mutant strain, H4 was still found to be acetylated (Mullen et al., 1989). yNatA or any other yNATs (yNatB and yNatC) require only a few specific amino acid residues for recognition and activity. However, the efficient acetylation by yNatD requires at least 30 to 50 residues at the N-terminus. It seems that during evolution, histones H4 and H2A have evolved to become the specific substrates of yNatD. In addition, yNatA can acetylate serine residue of modified H4 with changes at the N-terminus (Polevoda et al., 2009b). Besides histones H4 and H2A, no other proteins have been reported to be yeast NatD substrates.

1.3.5 Yeast NatE

Yeast NatE complex (yNatE) consists of the catalytic subunit yNaa50p (yNat5p) and the auxiliary subunit yNaa15p. yNatE is not well characterized. The catalytic subunit yNat5p, a homologue of catalytic subunits yArd1p, yNat3p and yMak3p, was discovered by Gautschi et al. (2003) during precipitating of yNatA complex. yNat5p was found in a complex with NatA by binding to auxiliary subunit Nat1p. However, the deletion of yNat5p did not display any visible phenotypes neither reduced the activity of NatA. Although no phenotype was observed in yeast *nat5*Δ*,* temperature sensitivity of two different double null mutant strains involved in ubiquitin-dependent protein degradation (*mck1 mds1* and *bul1 bul2*) was suppressed by the deletion of *yNAT5* (Andoh et al., 2000). Possibly, yNat5p is responsible for the N^{α}-acetylation of a small subset of proteins but so far none of yNatE substrates have been identified.

1.4 N-terminal acetylation in human

In human where 80-90% of soluble proteins are N-terminally acetylated all orthologous NATs from yeast (NatA-NatE) are found. Human and higher eukaryotes additionally express NatF (Van Damme et al., 2011c).

1.4.1 Human NatA

Human NatA (hNatA) complex is conserved from yeast regarding subunit homology and substrate specificity. hNatA consists of the catalytic subunit hNaa10p (hArd1p) and the auxiliary subunit hNaa15p (hNat1p/NATH) (Arnesen et al., 2005; Arnesen et al., 2009). However, the situation of NatA in human is more complicated than in yeast. Besides the conserved subunits hNaa10p and hNaa15p, two other paralogues of them hNaa11p (hArd2p) and hNaa16p (hNat2p) also participate to perform functions of NatA complexes (Arnesen et al., 2009). hNaa10p is homologous to the yeast NatA catalytic subunit yArd1p. It localizes to both the cytoplasm and the nucleus, and is found both in ribosome-bound form and non-ribosome-bound form (Arnesen at al., 2005). The auxiliary subunit hNaa15p localizes to the cytoplasm where it interacts with ribosome. However, a large fraction of hNaa10p and hNaa15p are not ribosome-associated (Arnesen at al., 2005). hNaa15p and hNaa16p are orthologous, they share 70% sequence identity. hNaa16p interacts with hNaa10p and it is found both in ribosome-bound form and non-ribosome-bound form (Arnesen et al., 2009).

Similar to yNatA, the hNatA complex activity requires the removal of methionine by methionine aminopeptidase. Once the methionine is removed, hNatA is able to acetylate Ser-, Ala-, Thr-, Gly-, Val-, and Cys- N-termini (Arnesen et al., 2009). Based on the substrate specificity and N-terminal sequences of all known human proteins (SwissProt v.56.0), it was estimated that more than 8000 human proteins are potential NatA substrates (Arnesen et al., 2009; Starheim et al., 2009b). Additionally, solely purified hNaa10p was reported to acetylate different subset of N-temini compared with NatA complex. Therefore, forming complex of NatA might alter the substrate specificity of catalytic subunit hNaa10p. Furthermore NatA or hNaa10p might function as a post-translational N^{α}-acetyltransferase (Van Damme et al., 2011b).

hNatA complex is a major NAT with more than 8000 potential substrates, its function indeed is important for the cell. Knockdown approaches have pointed out a role of hNatA in cell cycle and apoptosis. Depletion of NatA by siRNA caused growth inhibition, cell cycle arrest and induction of cell death in immortal human cell lines (Gromyko et al., 2010). In HeLa cells, knockdown of *hNAA10*, *hNAA15* and *hNAA16* induced apoptosis through the caspase dependent pathway, and possibly through G_o/G_1 cell cycle arrest (Arnesen et al., 2006; Arnesen et al., 2009). Furthermore, a point mutant (Ser37Pro) in hNaa10p, which resulted in the reduction of protein acetylation by NatA causes the lethal

X-linked disorder of infancy (Rope et al., 2011). However, many data show that the high expression of *hNAA10* and *hNAA15* is correlated with tumor development (Midorikawa et al., 2002).

1.4.2 Human NatB

The human NatB (hNatB) complex consists of the catalytic subunit hNaa20p (hNat3p) and the auxiliary subunit hNaa25p (hMdm20p) (Starheim et al., 2008; Ametzazurra et al., 2009). hNaa20p is homologous to yeast NatB catalytic subunit yNat3p and localizes to both cytoplasm and nucleus. hNaa25p, a homologue of yeast NatB auxiliary subunit yMdm20p, presents only in the cytoplasm. Both catalytic and auxiliary subunits of yNatB are found in ribosome-bond form and non-ribosome-bound form indicating that these proteins dynamically interact with ribosome or may have other functions (Starheim et al., 2008).

The substrate specificity of hNatB is not thoroughly identified. It was only reported that hNatB complex is able to acetylate peptide with a MDEL- N-terminus *in vitro* (Starheim et al., 2008). MDEL- is the N-terminus of NF-κB subunit p65, which is N-terminally acetylated by yNatB. Since yNatB acetylates methionine followed by an acidic residue, the acetylation of MDEL- peptide supports a theory of substrate specificity conservation of NatB from yeast to human (Polevoda et al., 2009a).

Similar to hNatA, knockdown of hNatB showed severe effects on cell growth, cell proliferation and cell cycle progression. Knockdown of *hNAA20* resulted in G_o/G_1 arrest , an increase in p21 level. Knockdown of *hNAA25* leaded to cell death, a decrease in G_o/G_1 cells and a decrease in p21 level (Starhiem et al., 2008; Ametzazurra et al., 2009).

1.4.3 Human NatC

The human NatC (hNatC) complex is recently identified. It composed of the catalytic subunit hNaa30p (hMak3p) and two auxiliary subunits hNaa35p (hMak10p) and hNaa38p (hMak31) (Starheim et al. 2009a). Like hNatA and hNatB, hNatC is conserved from yeast regarding subunit composition, localization and substrate specificity. As it was expected, all subunits of hNatC localize to the cytoplasm and stably associate with ribosomes (Starheim et al., 2009a). This localization supports the theory of hNatC to co-translationally acetylate nascent polypeptides like in yeast. However, the NatC subunits are

also found in the non-ribosome-bound form. Hence, they might dynamically associate with ribosomes or have some other functions that are independent of ribosome-bound form. Futhermore, hMak31p is also found in nucleus suggesting other functions of hMak31 beyond Nt-acetylation. It is possibly be true since hNaa31p is a member of the Sm and Sm-like proteins which associate with RNA and are involved RNA-processing event (Starheim et al., 2009a).

Similar to yeast, hNatC acetylate methionine followed by a hydrophobic amino acid residue. The catalytic subunit hMak3p *in vitro* N-terminally acetylates peptide with the N-termini MLALI-, MLGTG-, MLGTE- and MLGPE- (Starheim et al., 2009b). In addition, MLALI- is the terminus of hArl8b, a potential substrate of hNatC. The acetylation of methionine is essential for the lysosomal localization of hArl8b. In the *hMAK3* knockdown HeLa cells, the localization of HArl8b was aberrant (Starheim et al., 2009a; (Hofmann and Munro, 2006).

All of three subunits are essential for hNatC activity. Knockdown of each of the hNatC subunits resulted in the similar phenotype: reduced cell proliferation and induction of p53-dependent apoptosis (Starheim et al., 2009b).

1.4.4 Human NatD

Human NatD (hNatD), which contains only catalytic subunit hNaa40p (Nat4p) is recently described (Hole et al., 2011). The hNaa40p is conserved from yeast with 23% amino acid sequence identity. Besides sequence conservation, the hNaa40p is expected to co-translationally acetylate nascent polypeptides like its homologue yNaa40p. Indeed, both endogenous and exogenous hNaa40p are detected in both the cytoplasm and the nucleus. In addition, a minority of hNaa40p presents in the ribosome-bound form but most of hNaa40p is in a free form in the cytoplasm or in the nucleus (Hole et al., 2011). Therefore, hNaa40p might dynamically associate with ribosomes or has other ribosome-independent functions.

The substrate specificity of hNatA is highly conserved from yeast. hNaa40p N-terminally acetylates 177 unique substrates with preference for Ser-Gly-Gly-Gly-Lys starting peptides, which is very similar to histones H2A and H4 N-termini. Futhermore, hNaa40p certainly *in vitro* acetylate histones H2A and H4 but not histone H3. The efficient acetylation by hNaa40p requires peptides with only 7-19 amino acid residues length (Hole et al., 2011). This result contrasts with the finding of Polevoda et al. (2009b) which

postulated that acetylation by yNaa40p requires at least 30-50 amino acid residues at the N-terminus of Histones H4 and H2A.

Although *yNAA40* knockout mutant shows some minor phenotypes, the effects of knockout or knockdown function of *hNAA40* have not investigated yet.

1.4.5 Human NatE

The human NatE (hNatE) is the first NAT which is reported to have both N^{α}-acetyltransferase and N^{ε}-acetyltransferase (KAT) activities. Purified hNaa50p, a catalytic subunit of hNatE, acetylates methionine of oligopeptides with N-termini Met-Leu-. Besides, hNaa50p *in vitro* self-acetylates lysine residues 34, 37 and 142 of its own. In addition, histon 4 was detected as hNaa50p KAT substrate *in vitro* (Evjenth et al., 2009).

The catalytic subunit hNaa50p is the homologue of yeast Nat5p and the fruit-fly San protein (Gautschi et al., 2003; Williams et al., 2003). It was found in the cytoplasm in the associated form with hNaa10-hNaa15 complex (Arnesen et al., 2006). Thus, for a period of time, hNaa50p was thought to be the third subunit of hNatA complex. However, the knockout of *yNAA15* did not reduce the activity of yNatA (Gautschi et al., 2003). In contrast, knockdown of either *hNAA10* or *hNAA15* in HeLa cells caused a reduced level of hNaa50p (Hou et al., 2007). For these reasons, hNaa50p now is considered as catalytic subunit of a new NAT complex separately from hNatA but requires the whole hNatA as the auxiliary subunit (Evjenth et al., 2009; Starheim et al., 2009b).

1.4.6 Human NatF

Recently, a new type of N^{α}-acetyltransferase which particularly expresses in higher eukaryotes has been discovered and named NAT15/Q9H7X0/Naa60p (NatF). Naa60p is highly conserved among animals and its homologue also presents in plant (Van Damme et al., 2011c). Besides displaying distinct substrate specificity, hNaa60p N-terminally acetylates a larger subset of proteins. All of methionine residues followed by Lys-, Val-, Ala-, Ser- which so far are not identified as substrates of any known NATs are *in vitro* acetylated by hNaa60. Furthermore, hNaa60p also acetylates peptides with Met-Leu-, Met-Ile-, Met-Phe- at the N-termini which are believed as NatC substrates (Van Damme et al., 2011c). In addition, the *in vivo* identified substrates by expressing hNaa60p in yeast are in agreement with *in vitro* data set. hNaa60p is able to acetylates methionine residues of yeast

proteins followed by Lys-, Ser-, Val-, Leu-, Gln-, Ile-, Tyr- and Thr-. Statistically, the over-expression of hNaa60p increases the N^{α}-acetylation level of yeast proteins from 68% to 78% (Van Damme et al., 2011c).

Up to now, hNaa60p is the only one subunit of hNatF that has been identified but it is likely to have full functions as a N^{α}-acetyltransferase. Over-expression of *hNAA60* in Hela cells increases protein acetylation from 18% to 32%, whereas knockdown of *hNAA60* shows the opposite pattern (a decrease of acetylation from 26% to 17%). Furthermore, Naa60p functions are required for cell division. Depletion of *Drosophila NAA60* (*dNAA60*) resulted in chromosomal segregation defects during anaphase (Van Damme et al., 2011c).

1.5 N-terminal acetylation in plants

Although in yeast and human five and six NATs complexes have been identified one after another, the knowledge of Nt-acetylation in plants is surprisingly remaining scarce for many years. So far, in Arabidopsis, only four NAT complexes have been identified including NatC (Pesaresi et al., 2003), NatA, NatB and NatE (Stephan, 2011).

1.5.1 Identified NATs complexes

Arabidopsis NatC

Pam21p (At2g38130/AtMak3p/AtNaa20p) is the first N^{α}-acetyltransferase that has been described in plants (Pesaresi et al., 2003). It shares 60/68% identity/similarity to hNatC catalytic subunit hNaa20p and 52/62% identity/similarity to yNatC catalytic subunit yNaa20p. The *pam21* mutant had paler leaves and showed a reduction in growth rate and plant size. Further, knockout of *PAM21* reduced the synthesis of C1 and CP43 subunits PSII, resulting in a significant decrease in the abundance of PSII core proteins and of other thylakoid multiprotein complexes (Pesaresi et al., 2003).

The homologue of yMak10p, At2g11000 (AtMak10p), shares 32/25% similarity/identity amino acid sequence. AtMak10p likely interacts with AtMak3p according to the yeast hybrid assay (Pesaresi et al., 2003). However, Atmak3p is not required for the functions of AtNatC since *atmak10* mutant has no impact on the phenotype. In addition, At2g11000 protein cannot restore the function of yMak10p in *ymak10* strain indicating that At2g11000 does not act as a functional Mak10p homologue (Pesaresi et al., 2003).

Two other genes (*AT3G11500* and *AT2G23930*) encode Mak31p related proteins. In contrast to yeast, neither At3g11500 nor At2g23930 shows interaction with AtMak10p. Besides, no interactions between AtMak3p and At3g11500 or At2g23930 were detected (Pesaresi et al., 2003).

Taken together, it was suggested that AtMak3p alone can fulfill the entire functions of NatC. Indeed, AtMak3p was able to complement *ymak3p* and *ymak10* single mutants or *ymak3p ymak10* double mutant (Pesaresi et al., 2003).

Arabidopsis NatA complex

Arabidopsis NatA complex (AtNatA) and two other complex (AtNatB, AtNatC) were recently identified by Iwona Stephan in her PhD thesis (Stephan, 2011).

AtNatA is conserved from yeast and human with respect to protein sequence, subunit composition and substrate specificity. AtNatA consists of the catalytic sununit AtNaa10p (At5g13780, AtArd1p) and the auxiliary subunit AtNaa15p (At1g80410, AtNat1p). The catalytic subunit AtNaa10p displays high conservation to Naa10p from yeast and human with 44% and 67% amino acid sequence identity. However, the auxiliary subunit AtNaa15p shares only 30% and 41% amino acid sequence identity to yNaa15p and hNaa15p.

Similar to yeast and human, AtNaa10p was found to form a complex with AtNaa15p. Besides, both AtNaa10p and AtNaa15p localizes to the cytoplasm according to the co-localization with yellow fluorescent protein (YFP) indicating a co-translational N^{α}-acetyltransferase activity of AtNatA.

In addition, both catalytic subunit AtNaa10p and auxiliary subunit AtNaa15p are essential for AtNatA activity. Knockout of *AtNAA10* or *AtNAA15* both resulted in lethality of plants at the embryo globular state. Although both AtNaa10p and AtNaa15p subunits are needed for At NatA activity, solely purified AtNaa10p can *in vitro* N-terminally acetylate peptide with STPD- terminus. This result indicates that besides contributing to essential functions of AtNatA complex, AtNaa10p might act as an independent N^{α}-acetyltransferase which postranslationally acetylates a small subset of substrates like hNaa10p (Van Damme et al., 2011b).

Arabidopsis NatB complex

At1g03150 (AtNaa20p/AtNat3p) is the homologue of yeast and human NatB catalytic subunits with 40% and 58% amino acid sequence identity. The *atnaa20* knockout mutant

displays slower growth, small plant size and a reduction of N^{α}-acetyltransferase activity. In addition, 2-D gel and MS analyses of *atnaa20* proteome reveal three substrates of AtNatB: At4g09320 (nucleoside diphosphate kinase 1 protein), At1g13930 (unknown protein involved in salt stress) and At2g24940/MAPR2 (membrane-associated progesterone binding protein 2). The N-termini of At4g09320, At1g13930 and At2g24940 (Met-Asp- , Met-Asn-, Met-Glu-) suggest substrate specificity of NatB is conserved among species.

Although AtNaa20p is conserved from yeast regarding protein sequence and substrate specificity, yNaa20p cannot rescue the phenotype in *atnaa20* mutant. Surprisingly, AtNaa20p localizes not only to the cytoplasm but also to the mitochondria indicating the posibility of post-translational N^{α}-acetyltransferase activity of AtNaa20p in this organelle.

In addition, the homologue of yNatB and hNatB auxiliary subunit also presents in Arabidopsis (At5g58450) with 25% and 44% amino acid sequence identity respectively but so far it has not been heeded.

Arabidopsis NatE complex

The Arabidopsis NatE (AtNatE) catalytic subunit At5g11340 (AtNaa50p/AtNat5p) is conserved from yeast and human with 27% and 54% amino acid sequence identity respectively. According to the co-localization with YFP, AtNaa50p presents both in the cytoplasm and in the nucleus. Besides, AtNaa50p can *in vitro* acetylates peptide with Met-Leu- N-terminus like hNaa50p.

In yeast, the knockout mutant of *yNAA50* did not show any obvious phenotype (Gautschi et al., 2003). In contrast, *atnaa50* mutant displays severe affects including tiny plant, curve leaves, abnormal development and loss of ability to produce seeds. However, the overexpression of hNaa50p in *atnaa50* can rescue the wild type phenotypes.

1.5.2 Evidences of protein N^{α}-acetylation in plant organelles

For many years, it is believed that Nt-acetylation is a co-translational modification and only takes place in the cytoplasm. Recently, the new approach using mass spectrometry on proteomics researches has revealed the Nt-acetylation of protein in plant organelles. First, Kleffemann and coworkers reported 13 nuclear-encoded proteins that are N-terminally acetylated in rice chloroplasts after removal of the transit peptide indicating a post-translational modification in this organelle (Kleffmann et al., 2007). Later, in Arabidopsis chloroplasts, 4 chloroplast-encoded proteins and 47 mature nuclear-encoded proteins were

found with acetylated N-termini (Zybailov et al., 2008). Very recently, Bienvenut and coworkers have given for the first time an idea of the extent of Nt-acetylation in chloroplast with more than 220 characterized proteins (Bienvenut et al., 2011; Bienvenut et al., 2012). These results clearly highlight the high occurrence of this modification in the chloroplast compartment. Since these proteins are imported and processed to romove the transit peptide during their chloroplast translocation, it appears that the acetylation of these N-temini is a widespread protein post translational modification in chloroplast. Furthermore, the modified N-termini residues appear to be predominantly Ala, Ser, Val and Thr which are also preferential substrates of the conventional NatA (Bienvenut et al., 2012).

In addition, Sylvain Bischof in his PhD thesis has reported that Nt-acetylation occurs also in the mitochondrion where some mitochondrial-encoded proteins were found N-terminally acetylated (Bischof, 2010).

1.6 Aims of project

Nt-acetylation of protein is highly conserved among eukaryotic organisms. Up to now, five NAT complexes (NatA-NatE) in yeast and six NAT complexes (NatA-NatF) in human have been identified. In plants, however, only four complexes (NatA, NatB, NatC and NatE) have been reported as N^{α}-acetyltransferases, the knowledge of NatD and NatF in plants is still lacking. The first goal of project aim to identify and characterize these two missing NAT in *Arabidopsis thaliana*. Since NatD and NatF are believed to be conserved from yeast and human, the activities of putative AtNatD and AtNatF are tested using mimicked custom made peptides that are predictably acetylated by hNatD or hNatF. Subcellular localization reflects function of protein therefore the localization of AtNatD and AtNatF are addressed by fusing with EYFP and transiently expressed in Arabidopsis protoplasts. Furthermore, the biological functions of putative AtNatD and AtNatF are investigated by knockout function using T-DNA insertion lines.

Plants and other phototrophs differ from animals and yeast in the possession of chloroplasts. As mentioned in section 1.5.2, numerous chloroplast proteins are N-terminally acetylated but the knowledge about this process and its biological significances in chloroplast is vacant. The major aim of project is to identify and characterize the N^{α}-acetyltransferases responsible for the acetylation of chloroplast proteins. In order to identify substrate specificities of putative plastidic NATs, these proteins are expressed in *E. coli* and acetylated *E. coli* proteins are subsequently determined by LC-MS\MS. The biological roles of plastidic NATs are further characterized by knockout of their functions using T-DNA insertion lines.

2. MATERIAL AND METHODS

2.1 Technical equipment and materials

2.1.1 Technical equipment

Incubators (37°C/28°C)	Heraeus Instruments, Hanau
Electroporator MicroPulser	Bio-Rad, Munich
Photometer UvikonXL	Secoman, Kandsberg
Flow hoods: Lamin Air HB2448, HB2472	Heraeus Instruments, Osterode
Sonificator Sonoplus GM70	Bandelin Electronic, Berlin
Sonificator Tip UW 70	Bandelin Electronic, Berlin
Vortexer	Scientific Instruments, Ringoes (USA)
Table Centrifuge 5415 C	Eppendorf, Hamburg
Centrifuge 5417R	Eppendorf, Hamburg
Centrifuge J2-21, Rotor JA-20	Beckman, München
Centrifuge Sorvall RC5C, Rotor SCA-1500	DuPont, Bad Homburg
Centrifuge Rotanta 460R	Hettich GmbH, Tuttlingen, Germany
Ultracentrifuge Optima TLX-120	Beckman, München
Gel Jet Imager 2000	Intas, Göttingen
Plate Reader Fluostar Optima	BMG, Offenburg
Freezer (-80°C)	New Brunswick Scientific, Nürtingen
Mastercycler gradient 5531	Eppendorf
Mastercycler personal 5332	Eppendorf
Rotor-Gene Q	Qiagen, Hilden
Mini-Protean III electrophoresis and blot system	Bio-Rad, Munich
ImageQuant LAS 4000	GE Healthcare, München
Growth chambers	Waiss, Gießen
Liquid Scintillation Counter, Tri-Carb 2810 TR	ParkinElmer, USA
Cooling / Heating block Thermostat KBT-2 133	HLC, Bovenden
Heating block Thermostat HBT-2 132	HLC, Bovenden
Homogenisator Polytron®	Capitol Scientific
Autoclave Sanoklav	Sanoklav, Bad Überkingen- Hausen
Horizontal shaker The Belly Dancer	Stovall, Greensboro, NC (USA)
Microscope Leica DM IRB	Leica, Bensheim
Zeiss Confocal Microscope LSM510Meta	Zeiss, Jena

Further devices corresponded to the usual laboratory equipment.

2.1.2 Chemicals

[3H]Acetyl-CoA	Hartmann Analytic GmbH, Braunschweig
2-log DNA ladder	New England Biolabs, Beverly (USA)
4-Chlor-7-nitrobenzo-2-oxa-1,3-diazol	Sigma-Aldrich, Steinheim
Acetic acid	Merck, Darmstadt
Agar	Fluka Biochemika, Fuchs
Agarose	Serva, Heidelberg
Ampicillin	Roth, Karlsruhe
Bacto trypton	BD Biosciences, Heidelberg
Bacto yeast extract	BD Biosciences, Heidelberg
Bovine serum albumine	Sigma-Aldrich, Steinheim
Bromophenol blue	Feinchemie Kallies, Sebnitz
Calcium Chloride	Applichem, Karlsruhe
Coomassie Brilliant Blue G-250	Merck, Darmstadt or Bio-Rad, Hercules, CA (USA)
Deoxynucleotide Solution Mix (dNTP)	New England Biolabs, Beverly (USA)
D-Glucose	Merck, Darmstadt
Diethylpyrocarbonate (DEPC)	Roth, Karlsruhe
Dimethyl sulfoxide (DMSO)	Roth, Karlsruhe
Dithiothreitol (DTTred)	AppliChem, Karlsruhe
DNA loading buffer	Peqlab, Erlangen
Ethanol	Merck, Darmstadt
Ethidium bromide	Sigma-Aldrich, Steinheim
Ethylene diaminetetraacetic acid (EDTA)	Roth, Karlsruhe
Glycerol	Roth, Karlsruhe
Glycin	AppliChem GmbH, Darmstadt
HEPES	Roth, Karlsruhe
Imidazole	Sigma-Aldrich, Steinheim
Isopropanol	Roth, Karlsruhe
Isopropyl-D-1-thiogalactopyranoside (IPTG)	Duchefa, Haarlem (Netherlands)
Kanamycin	Duchefa, Haarlem (Netherlands)
Magnesium chloride	AppliChem, Karlsruhe

Magnesium sulfate	Merck, Darmstadt
Mannitol	Roth, Karlsruhe
MES	AppliChem, Karlsruhe
Methanol	VWR International GmbH, Darmstadt
Micro agar	Duchefa, Haarlem (Netherlands)
MOPS	AppliChem, Karlsruhe
Murashige Skoog (MS) incl. vitamins	Duchefa, Haarlem (Netherlands)
Nickelchlorid	AppliChem GmbH, Darmstadt
Phenylmethanesulphonylfluoride (PMSF)	Serva, Heidelberg
Polyethylenglycol (PEG) 4000	Roth, Karlsruhe
Potassium dihydrogenphosphate	Fluka Biochemika, Seelze
Potassium hydrogenphosphate	Fluka Biochemika, Seelze
Protease inhibitor mix	Sigma-Aldrich, Steinheim
Protein Standard Mark12TM	Invitrogen, Karlsruhe
PVDF membrane	AppliChem, Darmstadt
Rifampicillin	Duchefa, Haarlem (Netherlands)
Roti®-Quant Bradford reagent	Roth, Karlsruhe
Rotiphorese® Gel 30	Roth, Karlsruhe
Sodium azide	AppliChem, Darmstadt
Sodium carbonate	Grüssing GmbH
Sodium chloride	AppliChem, Darmstadt
Sodium dodecyl sulfate (SDS)	Fluka Biochemika, Seelze
SP Sephasore Beads	Sigma-Aldrich, Steinheim
Spectinomycin	Sigma-Aldrich, Steinheim
Sucrose, D+	AppliChem, Darmstadt
TEMED	Roth, Karlsruhe
Triton-X 100	Sigma-Aldrich, Steinheim
Tween-20	Sigma-Aldrich, Steinheim
Ultima Gold	PerkinElmer, USA
β-mercaptoethanol	Merck, Darmstadt

2.1.3 Consumables

96-well plate	Greiner, Frickenhausen
12-well plate	Corning GmbH, Kaiserslautern
Microscope Slides	Marienfeld, Laude-Königshofen

Microscope slides (cut edges)	VWR, Darmstadt
Counting chambers	Marienfeld GmbH, Lauda-Königshofe
HiTrap Chelating HP columns	GE Healthcare, München
Rotilabo aseptic filters (0,22 µm)	Roth, Karlsruhe
Rotilabo syringe filters (0.45 µM)	Roth, Karlsruhe
PVDF membrane	Millipore Corpotation, Billerica, USA
PD SpinTrap^{TM} G-25	GE Healthcare, München

2.1.4 Kits

NucleoSpin®Plasmid Kit	Macherey-Nagel, Düren
PCR Clean Up Kit®	Macherey-Nagel, Düren
QUIAGEN Plasmid Midi Kit	Qiagen, Hilden
RNeasy Plant Mini Kit	Qiagen, Hilden
cDNA synthesis Kit ReversAid® H-	Fermentas, Braunschweig
Rotor-Gene SYBR GREEN PCR Kit	Qiagen, Hilden

2.1.5 Enzymes and antibodies

Taq-DNA Polymerase (5 U/µl)	New England Biolabs, Beverly, (USA)
Phusion High fidelity DNA Polymerase	Finnzymes, Espoo (Finnland)
T4-DNA-Ligase	New England Biolabs, Beverly, (USA)
Gateway BP clonase II enzyme mix	Invitrogen, Karlsruhe
Gateway LR clonase II enzyme mix	Invitrogen, Karlsruhe
Cellulase „Onozuka R-10“	Serva, Heidelberg
Macerozyme R-10	Duchefa, Haarlem (Netherlands)
Restriction enzymes: *BamHI-HF^{TM}, KpnI-HF^{TM}, NcoI- HF^{TM}, SalI- HF^{TM}, EcoRV-HF^{TM}, XbaI.*	New England Biolabs, Beverly, (USA)
Acetylated-Lysine Antibody	New England Biolabs, Beverly, (USA)
Anti-Rabbit IgG HRP conjugated antibody	Promega Corporation, Madison, USA
Anti-GFP Antibody	HiSS Diagnostics, Freiburg, Germany
Anti-mouse IgG HRP-conjugated antibody	Sigma-Aldrich, Steinheim

2.1.6 Primers

Primers for Cloning

Number	Descreption	Sequence
1932	AT1G03650_for	GATCGGATCCATGGACAAAGGAGTTGTG
2084	AT1G03650_rev	GATCTCTAGAAATAGAGTCATCAAAATCTAGGTAC
2332	AT1G18335_for	GATCGGTACCATGGATCCGTCTCCTACCGA
2196	AT1G18335_rev	GATCTCTAGAATCTTTGGTAGGTTCTTCATCATT
1936	AT1G24040_for	GATCGGATCCATGGCGGCTTTAAGCATC
2133	AT1G24040_rev	GATCGTCGACCACATTTGCAGAGGAGGTCA
1938	AT1G26220_for	GATCGGATCCATGTTTCTCGGAGGCACA
2184	AT1G26220_rev	GATCGTCGACTTTCTTGTTTCTCTGTTTGCG
1940	AT1G32070_for	GATCGGATCCATGCTACTAATCCCAATTTC
2197	AT1G32070_rev	GATCGTCGACCTTTGGGTACCAAAACATGC
1942	AT4G28030_for	GATCGGTACCATGGCGTTTCTCTGTTCG
1943	AT4G28030_rev	GATCTCTAGACATGGACATGATTGGGGA
1944	AT4G19985_for	GATCGGATCCATGGGTTTGGTGGGTTGT
2333	At4g19985_rev	GATCTCTAGATGCCTCCAAGCTCTTTGTG
1946	AT2G02980_for	GATCGGATCCATGTCGCGATTTCCTCGA
2198	AT3G02980_rev	GATCGTCGACGACACAATCCGTGTCTTTGCA
1950	AT2G39000_for	GATCGGATCCATGCGGAGCACACCGTT
2186	AT2G39000_rev	GATCGTCGACCCGAAACTGTTCAAGAGCTTG
1952	AT2G06025_for	GATCGGTACCATGTCGACGATTTCGATC
1953	AT2G06025_rev	GATCTCTAGAGCTTGTGTACTGGAGCAAG
1954	AT1G72030_for	GATCGGATCCATGGGTCACTTGCCACA
2134	AT1G72030_rev	GATCTCTAGAAGAAAAGCGTTTACTCATAAGAAC
1956	AT5G16800_for	GATCGGATCCATGTCGCGTTTTCCCC
2085	AT5G16800_rev	GATCGTCGACTACGTAATCATACCCGGATGA
2367	E41_At5g16800_f	GATCCCATGGACATGTCGCGTTTTCCCCGT
2339	E41_At5g16800_r	GATCGGATCCCTATACGTAATCATACCCGGATG
2392	E41_At1g18335_f	GATCCCATGGACATGGATCCGTCTCCTACCGA
2393	E41_At1g18335_r	GATCGGTACCTCAATCTTTGGTAGGTTCTTCATC
2578	E41_At1g24040_f	CCATGGACACAGCTACAGAAACTGGAGAAGAAA
2579	E41_At1g24040_r	GGATCCTTACACATTTGCAGAGGAGGTC
2588	E41_At2g39000_f	CCATGGACGCAAGTCAAATAGTTGATCTTTTTCC
2589	E41_At2g39000_r	GGATCCTTACCGAAACTGTTCAAGAGCTTG

2590	E41_At2g06025_f	CCATGGACAGTCACTGGGAAGATCGCTCC
2591	E41_At2g06025_r	GGATCCTTAGCTTGTGTACTGGAGCAAGTATG
3125	GW_Naa60_for	GGGGACAAGTTTGTACAAAAAAGCAGGCTgcATGTCG CGTTTTCCCCGT
3126	GW_Naa60_revC	GGGGACCACTTTGTACAAGAAAGCTGGGTcTACGTAA TCATACCCGGATGAT
3127	GW_Naa60_revN	GGGGACCACTTTGTACAAGAAAGCTGGGTcCTATACG TAATCATACCCGGATG

Primers for genotyping

Number	Descreption	Sequence
2425	Naa60_LP	TTACCCAAGAAACAGCCATTG
2426	Naa60_RP	TGATTCTGGAAACCCTAACCC
2499	Naa40.1_LP	ACAATCTTTGCAACCGATTTG
2500	Naa40.1_RP	TGCATGAGTAAGGGATTCACAG
2501	Naa40.2_LP	TTTGACCCGTCATGTTCTTTC
2502	Naa40.2_RP	TGTGATGTTTCACAGACGAGG
2723	N306363_LP	GAAGGAGGAGAAGCATTAGCG
2724	N306363_RP	CATGACGATCCCTCTGAACTC
2725	N553123_LP	AGCCACGAGATTCTTCTCTCC
2726	N553123_RP	TGAGTGGAGATGATTGGTATGTG
2727	N572318_LP	CAGATTTCTACGTCACATTGGTG
2728	N572318_RP	CAGTGACAAGAGGTCCTCAGG
2787	N532239_LP	TTTTGACCACAAAAATTGCAG
2788	N532239_RP	ATTCGAAGATTTGGGTCTTCG
2793	N678583_LP	GCAAGAAAGAATGCAGCAAAC
2794	N678583_RP	AACCATTCCTTTGATCCCATC
2814	N849101_LP	TTGTGGCCTATTGAGGTATCG
2815	N849101_RP	TATCCGAATCGGTGTCAGAAG
432	GA_LB1	CCCATTTGGACATGAATGTAGACAC
1401	LB1.3new	ATTTTGCCGATTTCGGAAC
789	Wisc_LB_p745	AACGTCCGCAATGTGTTATTAAGTTGTC

Primers for qRT-PCR

Number	Descreption	Sequence
2142	Actin7_for_new	CCGTTCTTTCTCTCTATGCCAGTG G
2143	Actin7_rev_new	GGATAGCATGAGGAAGAGCATACCC
2459	PP2a_PDF2_for	CTTCTCGCTCCAGTAATGGGATCC
2460	PP2a_PDF2_rev	GCTTGGTCGATCATCGGAATGAGAG
2826	DREB2A_fwd	ATGAAAGGTAAAGGAGGACCAG
2827	DREB2A_rev	CAAAGCCTGCTACCTCGATTA
2569	qRT_naa40.1_for	CGGCGAAAGATTTTGGAGAAG
2579	qRT_naa40.1_rev	TCCACGGCCAGATTCCAGGTAC
2663	qRT_naa40.2_for	CTAAGAATCGGATGAGTGCGATCG
2664	qRT_naa40.2_rev	CACCGAGAGAGTCGGAAGATTAGC
2573	qRT_At5g16800_for	ACGTGGAATCGCTAAGGCAC
2574	qRT_At5g16800_rev	TTTTGTACAACCGGATTGCGGG
2797	qPCR_At1g32070_For	GCAGCAGGAGAAGAAGCTTATTGG
2798	qPCR_At1g32070_Rev	CCTAGCCCTTGACCCTGATATTC
2799	qPCR_At1g26220_For	TGATGGAGCGGTTGATTGAG
2780	qPCR_At1g26220_Rev	GTCAGAGACGAACCCAAGCG
2920	qPCR_At2g06025_for	GCCAGACTGAGTGGAGTTGAACA
2921	qPCR_At2g06025_rev	ACCGGTCTTCTGGTATAATTCCTG
2922	qPCR_At2g39000_for	CAAGATCCCTGAAGGAGCGAC
2923	qPCR_At2g39000_rev	CCGAAACTGTTCAAGAGCTTGTG
2924	qPCR_At4g28030_for	CCAGAGGGGATTGGTGTGGA
2925	qPCR_At4g28030_rev	CCCCATTCTCCAGCAACTCTC
3113	qPCR_At1g24040_For	CTCTCAAACCAACCCAATCAACG
3114	qPCR_At1g24040_Rev	GGAGGGAAAGCGAATCTGAG
3123	qPCR_At4g19985_for	GTCAGGCTACTCACAAGCAGAGAC
3124	qPCR_At4g19985_rev	CCTCGAAGCAGAGAGCTGCT

Primers for sequencing

Number	Descreption	Sequence
334	M13 r	TTCACACAGGAAACAG
335	M13 u	GTAAAACGACGGCCAGT
689	pDONR201-Seq-For	TCGCGTTAACGCTAGCATGGATCTC

690	pDONR201-Seq-Rev	GTAACATCAGAGATTTTGAGACAC

2.1.7 Softwares

Zeiss LSM Image Software	Zeiss, Jena
SigmaPlot 12.0	SPSS Inc., Munich
Rotor-Gene® Q Series Software	Qiagen, Hilden
Photoshop CS 7.0.1	Adobe Systems GmbH, Munich
Fluostar Optima 1.30	BMG Labtechnologies, Offenburg
EndNote X2	Thomson Reuters, New York, NY (USA)
ImageQuant TL	GE Healthcare, München
Vector NTI 9	Invitrogen, Karlsruhe
Microsoft Office	Microsoft (USA)
TreeView 1.6.6	Taxonomy and systematics, Glasgow

Web based software tools and websites:

Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/
TAIR	www.arabidopsis.org
Primer calc	www.basic.northwestern.edu/biotools/oligocalc.html
PredictProtein	https://www.predictprotein.org/
TargetP1.1	http://www.cbs.dtu.dk/services/TargetP/
TM finder	http://tmfinder.research.sickkids.ca/cgi-bin/TMFinderForm.cgi
Smart	http://smart.embl-heidelberg.de/
NSL-mapper	http://nls-mapper.iab.keio.ac.jp/
NLStradamus	http://www.moseslab.csb.utoronto.ca/NLStradamus/

2.1.8 Buffers and solutions

Enzyme solution	W5 solution	PEG solution
1% cellulase 0.25% macerozyme 0.4 M mannitol 10 mM $CaCl_2$ 20 mM KCl 0.1% BSA 20 mM MES pH 5.7	154 mM NaCl 125 mM $CaCl_2$ 5 mM KCl 5 mM glucose 2 mM MES pH 5.7	30% (w/v) PEG 4000 0.1 M $CaCl_2$ 0.2 M mannitol No pH adjusted
MMG solution	**WI solution**	**Edward's buffer**
0.4 M mannitol 15 mM $MgCl_2$ 4 mM MES pH 5.7	4 mM MES 0.5 M mannitol 20 mM KCl pH 5.7	200 mM Tris-HCl pH 7.5 250 mM NaCl 25 mM EDTA 0.5% SDS
Buffer B	**Buffer W**	**Buffer E**
50 mM Tris (pH 8.0) 250 mM NaCl 20 mM Imidazol	50 mM Tris (pH 8.0) 250 mM NaCl 100 mM/150 mM Imidazol	50 mM Tris (pH 8.0) 250 mM NaCl 400 mM Imidazol
Acetylation buffer	**KAT buffer**	**DNA gel loading buffer**
500 mM Tris/HCl pH 7.5 10 mM DTT 8 mM EDTA	50 mM Tris/HCl pH 8 10% glycerol 1 mM EDTA	0.25% bromophenol blue 0.25% xylene cyanole 40% glycerol
LB media	**Plant protein extraction buffer**	**Laemmli buffer (5x)**
1% (w/v) tryptone, pH 7 0.5% (w/v) yeast extract 1% (w/v) NaCl for solid media: 1.5% (w/v) agar	50 mM Hepes pH 7.4 10 mM KCl 1 mM EDTA 1 mM EGTA 10% (v/v) glycerol	100 mM Tris, pH 7.0 20% (v/v) glycerine 10% (w/v) SDS 0.1% bromphenolbllue 25% β-mercaptoethanol

Blotting buffer	SDS-PAGE running buffer (10x)	Tris Boric acid EDTA (TBE) buffer
40 mM Tris, pH 8.0 190 mM glycin 0.1% (w/v) SDS 20% (v/v) methanol	250 mM Tris 1.92 M glycin 1% (w/v) SDS no pH adjustments	90 mM Tris-HCl pH 8.0 90 mM boric acid 0.5 mM EDTA
Tris-buffered saline (TBS)	**Phosphate buffered Saline (PBS)**	**Plasma membrane lysis buffer**
20 mM Tris-HCl, pH 7.6 137 mM NaCl	NaCl 137 mM, pH 7.4 KCl 2.7 mM Na_2HPO_4 10 mM KH_2PO_4 1.8 mM	200 mM sorbitol 10% Fircoll 20 mM EDTA 10 mM HEPES

2.1.9 Oligo peptides

Name	Sequence
Histon H4-ter	H-SGRGKGGKGR-RWGRPVGRRRRPVRVY-OH
AKR1-ter	H-MVNALEN- RWGRPVGRRRRPVRVYP-OH
AASS-ter	H-AASSVGN- RWGRPVGRRRRPVRVYP-OH

All peptides were ordered from JPT Peptide Technologies, Berlin, Germany.

2.2 Plant materials and methods

2.2.1 Plant materials

- *Arabidopsis thaliana* ecotype Col-0 (family: *Brassicaceae*, order: *Capparales*, class: *Dicotyledonae*, subdivision: *Angiospermae*).

- T-DNA insertion lines obtained from NASC (Table 2.1)

- *Nicotiana benthamiana* (family: *Solanaceae,* order: *Solanales,* class: *Dictyledoneae,* subdivision: *Angiospermae*).

Table 2.1 List of T-DNA insertion mutants were used in this work.

GMO number	T-DNA insertion lines	Locus	Description
1060	SALK_152065C	At1g18335	*naa40-1*
1061	SALK_081588	At1g18335	*naa40-2*
1121	SALK_016406C	At5g16800	*naa60*
1123	GK-034E09.02	At1g24040	*pnaa20*
1124	GK-809D05	At4g19985	*Putative NAT*
1125	SALK_032239	At1g32070	*Putative NAT*
1126	SALK_053123	At2g06025	*pnaa30*
1127	SALK_072318	At2g39000	*pnaa10*
1132	SALK_062388C	At1g26220	*Putative NAT*
1134	WiscDsLox4B08	At4g28030	*Putative NAT*

2.2.2 Growth conditions

Arabidopsis seeds were planted on humid soil (1 Vermiculite : 2 sand : 7 ED 73) in the pots and stratified at 4^{o}C in the dark. After 3 days, the seeds-planted pots were transferred to short day climate chamber (8.5 h/15.5 h day/night cycle; 100 µE m-2s-1; 22°C day/18°C night; 50 % humidity). 2 week old seedlings then were separated into new pots and kept growing under short day conditions. When plants were old enough (about 7-8 week old), they were transferred again to the long day chamber (16 h/8 h day/night cycle; 130 µE m-2s-1; 22°C day/18°C night; 50 % humidity) for producing seeds.

2.2.3 *In vitro* growth of Arabidopsis seedlings on agar plates

The seeds were sterilized as described by (Meyer and Fricker, 2000) and grown on 0.8% (w/v) agar plates supplemented with Murashige & Skoog (MS). For further experiments, other reagents (NaCl, mannitol, glucose…) were added to the medium. The medium was

autoclaved, cooled down at about 60°C and then poured in the plates. Seeds on the plates were stratified at 4°C for 2 days before transferring to the short day conditions.

2.2.4 Drought stress treatment

The seeds were planted on soil as described above (2.2.2). After growing normally at the short day conditions for 4 weeks, plants were stopped watering for 12-14 days, whereas the control plants were watered regularly.

2.2.5 UV-c tress treatment

UV stress treatment was carried out as described by (Brown et al., 2005), UV-c was used instead of UV-b. In brief, three week-old Arabidopsis plants were exposed with UV-c in the sterile hood Lamin Air HB2472 for different periods of time. After UV-c exposing, the plants were recovered under the short day conditions.

2.2.6 Arabidopsis protoplasts isolation

Arabidopsis protoplasts were isolated as described by (Wu et al., 2009) with some minor modifications. Leaves were collected from 4-5 week-old plants grown under short day conditions. The upper epidermal surface was stabilized b affixing with a strip of Time tape (Time Med, Burr Ridge, IL), whereas the lower epidermal surface was fixed to a strip of Magic tape (3 M, St. Paul, MN). The Magic tape was then carefully pulled away from the Time tape, peeling away the lower epidermal surface cell layer. The peeled leaves (15-20 leaves), still adhering to the Time tape, were transferred to a Petri dish containing 10 ml of enzyme solution. The leaves were gently shaken (40 rpm on a platform shaker) in light for 60 min until the protoplasts were released into the solution. The protoplasts were mixed 1 to 1 with pre-chilled W5 and centrifuged at 100g for 2 min (swing out rotor) then they were combined and washed with 50 ml of W5 solution and incubated on ice for 60 min. During the incubation period, protoplasts were counted using a hemocytometer (Fuchs-Rosenthal) under light microscope. W5 solution was removed without touching the protoplasts pellet. The protoplasts pallet then was re-suspended at 2.10^5 ml^{-1} in MMG solution and kept at room temperature.

2.2.7 PEG-Calcium transfection of Arabidopsis protoplasts

Transfection of Arabidopsis protoplast was done as described by (Yoo et al., 2007) with some minor modifications. 20 µl of plasmid DNA (10-20 µg) was mixed gently with 100 µl protoplasts then 120 µl PEG solution was added. The mixture was mixed again gently and incubated at room temperature for 5 min. 500 µl W5 solution was added to stop the transfection, next the samples were centrifuged at 100g for 2 min. Supernatant was removed, transformed protoplasts were carefully re-suspended in 500 µl WI solution and gently transferred to 12 well plates (coated with 5 % sterile BSA). After incubation overnight in the dark at the room temperature, protoplasts were analyzed using confocal microscopy.

2.2.8 Osmotic plasma membrane lysis of protoplasts

To release the vacuoles protoplasts, solution was mixed with osmotic lysis buffer (section 2.1.8) on the slice right before analysis using confocal microscopy.

2.2.9 Infiltration of tobacco leaves

One colony of *Agrobacterium tumefaciens* (strain GV3101) carrying binary vector was grown in 5 ml of LB medium supplemented with appropriate antibiotics (25 µg/ml Rif, 30 µg/ml Gent and 50 µg/ml Strep) overnight (16-18h) at 28^{o}C, shaking 200 rpm. The overnight culture was diluted 1/10 with fresh LB medium containing antibiotics and grown at 28^{o}C, shaking 200 rpm to get OD_{600nm} 0.8-1 (about 3-4h). Bacteria were pelleted by centrifuging for 10 min, 5000 rpm at 4^{o}C then washed twice with ice cold sterile water. Finally, bacteria were re-suspended with ice cold water to get an OD_{600nm} 0.8. *Nicotiana benthamiana* plants were kept 100% humidity for at least 1 hour before infiltration. The bacteria were infiltrated into the leaves using 1ml syringe without a needle. The expression of transgene was analyzed 2 days after infiltration.

2.3 Bacteriological Methods

2.3.1 Bacterial strains

- *E. coli* XL1-blue

- *E. coli* Rosetta

- *Agrobacterium tumefaciens* GV3101

2.3.2 Bacterial growth

Bacteria were grown in the LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7). For plates, media was supplemented with 1.5% agar before autoclaving. Transformed bacteria were selected on the LB plates containing appropriate antibiotics with following concentrations: ampicillin 100 μg/ml, kanamycin 50 μg/ml, rifampicin 25 μg/l, spectinomycin 50 μg/ml, gentamycin 30 μg/ml. *E.coli* strains were grown at 37^{o}C, whereas *Agrobacterium tumefaciens* was grown at 28^{o}C.

2.3.3 Preparation of electro-competent cells

E. coli strains XL1-blue and Rosetta were used to prepare electro-competent cells. Bacteria were grown in 25 ml LB medium overnight. 10ml of overnight culture was transferred to 250 ml LB and grown to get OD_{600nm} 0.8 (about 3h). The culture was cooled down on ice for 30 min then pelleted by centrifuging at 2500 rpm for 10 min at 4^{o}C. Next, the pellet was washed twice with 250 ml and 125 ml ice-cold water followed by 2 other washing steps with 125 ml and 50 ml 10% ice-cold glycerol. Finally, pellet was re-suspended with 2.5 ml 10% glycerol. Aliquots of 40 μl cell solution were transferred to cold Eppendorf tubes, frozen in liquid nitrogen and stored at -80^{o}C until use.

2.3.4 Preparation of Agrobacterium competent cells

Agrobacterium tumefaciens strain GV3101 was grown in 5 ml LB medium for 24 hours at 28^{o}C. 4ml of overnight culture was transferred to 200 ml of LB medium (supplemented with Rif and Gent) and grown at 28^{o}C to get an OD_{600nm} 0.5. Cells were pelleted by centrifuging for 10 min at 5000 rpm. Supernatant was discarded and the pellet was re-suspended in 50 ml NaCl 150 mM. After centrifuging again for 10 min at 5000 rpm, cells were re-suspended in 10 ml ice cold $CaCl_2$ 20 mM. An aliquot of 200 μl competent cells was transferred to the ice cold tube and stored at -80^{o}C until use.

2.3.5 Electro-transformation of bacteria

The competent cells (section 2.3.3) were transformed by electroporation at 2500V and 15μF and incubated in 1 ml LB medium for one hour at 37°C, shaking at 220 rpm. 100 μl of bacterial suspension was spread on solid LB plate containing selective antibiotic and grown overnight at 37°C.

2.3.6 Heat-shock transformation of Agrobacterium

The competent cells (section 2.3.4) were thawed on ice for 1 hour, mixed with 1μg plasmid DNA and incubated on ice for 10 min. Cells were then frozen in liquid nitrogen for 5 min following by heat shocked at 37°C for 5 min. After adding 800 μl LB medium, the cells were recovered at 28°C for 4 hours. Finally, cells were spread on selective LB medium and grown at 28°C for 2-3 days.

2.3.7 Plasmids isolation of bacteria

The bacteria carrying plasmids were grown overnight in LB medium with selective antibiotic. Plasmid mini-preparation/midi-preparation was done with NucleoSpin Plasmid Kit®/QUIAGEN Plasmid Midi Kit according to the manufacturer's protocol.

2.3.8 Bacterial glycerol stock

Bacteria were grown in 4 ml LB medium supplemented with selective antibiotic overnight. 800 μl of overnight culture was transferred to Eppendorf tube containing 200 μl 80% glycerol. The mixture was then frozen in liquid nitrogen and stored at -80^{o}C.

2.3.9 Total protein extraction from *E. coli*

1 ml cell culture was pelleted by centrifuging 10 000 rpm for 2 minutes. The sediment was resuspended in 150 μl of 1x Leammli buffer and boild at 95^{o}C for 5 min. 10 μl of mixture was used for SDS-PAGE.

2.4 Molecular Biology Methods

2.4.1 Genomic DNA extraction from *Arabidopsis thaliana*

Genomic DNA was extracted as described by Edwards et al., (1991). Leaves were ground in a microtube with plastic pestle in 400 µl Edward's buffer. The samples were vortexed for 10s and centrifuged at 13000 rpm for 5 min. 300 µl supernatant was transferred to the new tube, an equal volume of isopropanol was added and mixed gently by inverting the tube. After incubating for 2 min at room temperature, genomic DNA was participated by centrifuging at 13000 rpm for 10 min. The supernatant was discarded, the pellet was washed with 700 µl Ethanol 70% and dried at RT for 15 min. 50 µl sterile water was used to resolve the DNA. 1 µl of DNA preparation was used for PCR.

2.4.2 Polymerase chain reaction (PCR)

Specific DNA fragments were amplified by PCR using MasterCycler. Two different PCR protocols were applied, one for genotyping and one for cloning. The PCR components and programs were set up according to manufacturer's protocol. The optimal temperature for primers binding was calculated using Oligo Cal (2.1.7).

PCR components for cloning	#	Step	Temperature	time
5XBuffer: 10 µl	1	initial denaturing	98	1 min
10 mM dNTPs: 1 µl	2	denaturing	98	15 sec
10 µM primer F: 1µl	3	annealing	Primer specific	20 sec
10 µM primer R: 1 µl	4	elongation	72	15 sec
Template DNA: 1 µl	5	final elongation	72	7 min
Phusion DNA polymerase: 0.5 µl	Steps 2-4 were repeated 35 times			
H_20: up to 50 µl				

PCR components for genotyping	#	Step	Temperature	time
10X Buffer: 2.5 µl	1	initial denaturing	95	5 min
10 mM dNTPs: 0.5 µl	2	denaturing	95	30 sec
10 µM primer F: 0.5 µl	3	annealing	Primer specific	30 sec
10 µM primer R: 0.5 µl	4	elongation	72	1 min
Template DNA: 1 µl	5	final elongation	72	7 min
Taq DNA polymerase: 0.125 µl	Steps 2-4 were repeated 35 times			
H_20: up to 25 µl				

2.4.3 DNA gel electrophoresis

DNA fragments were separated by electrophoresis in 0.8% – 2% agarose gels. For gel preparation, agarose was melted in 1X TBE buffer, cooled down to about 60^{o}C and ethidium bromide was added to final concentration 0.5 µg/ml. DNA was mixed with loading buffer before loading then separated by applying a current of 120V in TBE running buffer. Finally, gel was documented by GelJet Imager 2000.

2.4.4 Cloning using Endonucleases and Ligase

Fragment amplification

Specific fragment for cloning was amplified using Phusion High Fidelity DNA Polymerase and PCR as described in (section 2.4.2).

RE digestion of DNA

10 µl of plasmid DNA was mixed with 10 U of restriction enzymes in suitable buffer with a total volume of 100 µl. The reaction was incubated overnight at 37^{o}C to open plasmids or cut out the fragments.

DNA purification from agarose gel

After separating on agarose gel, the fragments of PCR products or open plasmids were cut out and purified using PCR Clean Up Kit® (section 2.1.4) according to the manufacturer's protocol. The concentration of purified DNAs were measured using NanoDrop 2000 spectrophotometer (section 2.1.1)

DNA ligation and transformation

The ligation reactions were done with the total volume of 20 µl (1 µl T4ligase with buffer). For blunt end ligation, PCR products were mixed in a ratio 10:1 with pBSSK and incubated for 2 hours at RT. For sticky end, the insertion and plasmid were mixed in a ratio 3:1 and incubated for 15 min. Amounts of DNA were calculated according to the formula (ng vector/bp vector) = (ng insert/bp insert). 1 µl of the ligation was used for transformation into *E. coli* XL1-blue (section 2.3.5).

DNA sequencing

For sequencing, 1µg-1.5µg plasmid DNA with primer was sent to the company. The sequencing was done by StarSeq GmbH, Mainz, Germany or Eurofins MWG GmbH, Ebersberg, Germany. Sequencing results were checked using Vector NTI 9 (section 2.1.7).

Constructs created by using endonucleases and ligase were listed in the Table S2.

2.4.5 Gateway cloning

Primer design and all procedures were done according to the manufacturer's protocols. Created gateway constructs were listed in the Table S2.

2.4.6 Total RNA extraction from *Arabidopsis thaliana*

Approximately 100mg of leaves was used for total RNA extraction using RNeasy Plant Mini Kit according to the manufacturer's protocols. RNA concentration was determined using NanoDrop 2000 spectrophotometer (section 2.1.1).

2.4.7 cDNA synthesis

cDNA was synthesized from 2µg of total RNA extract using RevertAid™ H Minus First Strand cDNA Synthesis Kit (section 2.1.4) according to the manufacturer's protocols.

2.4.8 qRT-PCR

The 1/10 diluted cDNA was used for qRT-PCR. The reaction was performed following component (2 µl cDNA, 6.25 µl SYBR mix, 0.5 µl primers (5µM each) and 3.75 µl H_20) and program:

Step 1: 95 °C 5 min
Step 2: 95 °C 5 sec
Step 3: 60 °C 10 sec ; acquiring to Cycling A ([Green][1][1])
Melt: (60-95 °C), hold 5 secs on 1st step, hold 5 secs on next step, Melt A([Green][1][1])
(Step 2-3 were repeated 40 times)
Rotor-Gene® Q Series Software (section 2.1.7) was used for quantification.

2.5 Protein methods

2.5.1 Total protein extraction from *Arabidopsis thaliana*

For protein extraction 200mg leaf material was ground in liquid nitrogen. 500 µl extraction buffer (section 2.1.8) (supplemented with 10 mM DTT, 0.5 PMSF) was added, vortexed and incubated on ice for 15 min. The samples were centrifuged at 14000 rpm for 10 min at 4^{o}C. The supernatant was collected and used for further experiments.

2.5.2 Purification of total protein extract

The total protein extract (section 2.5.1) was purified using PD SpinTrap^{TM} G-25 column (section 2.1.3) according to the manufacturer's protocols.

2.5.3 His-tag protein purification

E. coli Resetta carrying expression vector was grown in LB medium (supplemented with selection antibiotic) and induced by IPTG (final concentration 1mM/l) at 37 ^{o}C for 5 hour or at 21^{o}C overnight. Cells were pelleted by centrifuged at 6000 rpm for 5 min at 4^{o}C. The pellet was re-suspended in 6ml buffer B (section 2.1.8) and 25 µl 200 mM PMSF, sonicated twicw on ice for 5. After centrifuging at 15000 rpm for 10 min at 4^{o}C, the supernatant was circulating for 1 hour in HiTrap^{TM} Chelating column (section 2.1.3). The column was washed with buffer W (section 2.1.8) for 10 min and finally His-tag protein was eluted with buffer E (section 2.1.8).

2.5.4 *In vitro* Acetyltransferase assay

Purified enzyme (overnight induction) was mixed with oligopeptide (section 2.1.9), ^{3}H-Acetyl CoA, BSA and acetylation buffer (section 2.1.8) supplemented with 10 mM DTT in the total volume of 100 µl. The mixture was incubated at 37^{o}C for different periods of time. For different experiments, different concentrations of reagents of reaction were used. After incubation, samples were centrifuged at 1500g for 4 min. The supernatant was mix with 100 µl SP Sepharose (50% in 0.5M acetic acid) and incubated in the rotor for 5 min. After centrifuging for 4 min at 1500g, the pellet was washed 3 times with 0.5M acetic acid and 1 time with 100% methanol. The radioactive was measure by Scintillation counter (section 2.1.1).

2.5.5 *In vitro* Auto Lysine acetylation assay

For Auto-acetylation assay 20 µg purified MBP-pNat1 (5 hours induction) was mixed with 5 µl Acetyl-CoA 2 mM and KAT buffer (section 2.1.8) in the total volume of 80 µl. The mixture was incubated at 37^{o}C for 10-60 min. Reaction was stop by adding 20 µl 5X Laemmli buffer (section 2.1.8). The amount of 100 ng protein was loaded on the gel for Western Blot detection.

2.5.6 NBD-Cl fluorescence assay

NBD-Cl fluorescence assay was done as described by (Bernal-Perez et al., 2012) with minor modifications. 7.6 µg purified proteins (section 2.5.2) was mixed with 0.5 mM NBD–Cl in 50 mM sodium citrate buffer containing 1 mM EDTA (pH 7.0) and incubated in the dark overnight at RT. The solution was excited at 470 nm and fluorescence emission intensity was recorded at 520 nm using Plate Reader Fluostar Optima (section 2.1.1).

2.5.7 Preparation and analysis of membrane fractions

200mg leaf material was ground in liquid nitrogen and 600 µl extraction buffer (section 2.1.8) (supplemented with protease inhibitors mix ,10 mM DTT and 0.5 mM PMSF). After incubating on ice for 15 min, the samples were centrifuged 2 times at 20 000g for 5 min to get rid of cell debris, nuclei, ER etc. 1ml of clear supernatant (combined from 2 tubes) was transferred to 1ml-ultracentrifuge tube and centrifuged at 100 000g for 1 hour at 4^{o}C. The supernatant (S1) and pellet (M1) were carefully separated. The pellet (M1) was either re-suspended in 150 µl PBS (containing protease inhibitors) or 600 µl 0.5 M Na_2CO_3. Na_2CO_3 is able to disrupt the protein interaction (Marmagne et al. 2007). The Na_2CO_3 treated sampled were incubated on ice with frequently mixing for 30 min and then centrifuged again at 100 000g for 1 hour at 4^{o}C. The pellet (M2) was re-suspended with 100 µl PBS (containing protease inhibitors). The supernatant (S2) was transferred to Eppendorf tube and proteins were participated by TCA. TCA participated pallet was re-suspended in 100 µl 2x Laemmli buffer and heated up at 95^{o}C for 10 min before loading on gel.

2.5.8 Protein SDS-PAGE, Western Blot and Immunodetection

For SDS-PAGE proteins were mixed with 5X Laemmli buffer to get final 1X buffer concentration before loading on 12.5% polyacrylamide gel. Proteins were separated by applying a current of 80 V for 10 min and following 120 V for 2 hours in protein running buffer (section 2.1.8). The proteins then were bloted from gel into PMVD membrane by applying a current of 260 mA for 2.5 hours in blotting buffer (section 2.1.8) at 4^{o}C. After blotting the membrane was blocked with 5% BSA in 1X TBS buffer for 1 hour. Blocked membrane was washed 5 times with 1X TBS and then incubated with the first antibody (1:10 000) overnight at 4^{o}C. The next day, membrane was washes again 5 times before incubating with the second antibody (1:20 000) for 1.5 hours at RT. The membrane then was washed, incubated with 1 ml working solution of SuperSignal West Dura Extended Duration Substrate (ThermoScientific) for 5 min in plastic foil. Finally, the signals were developed and detected using ImageQuant LAS 4000 (section 2.1.1). The images were analyzed using ImageQuant TL (section 2.1.8).

After detection, the PMVD membrane was stained with Amido black Staining Solution (2.1.8) to detect other bands and marker.

2.5.9 LC-MS/MS analysis

For LC-MS/MS analysis, the pellets and crude extracts of *E. coli* Rosetta that express putative pNATs were sent to proteomics collaboration partner. All steps of proteins preparation and LC-MS/LS analysis were done by Willy Bienvenutin the lab of Carmela Giglione, Insitut des Sciences du Vegetal, CNRS, France. In brief, 2 mg of proteins of each sample was subject of reduction and Cys alkylation. Then, the free amino groups (Protein free N-termini and e-NH_2 of the Lys) were subject of D3-acetylation followed by trypsin digestion. The digested mixture was fractionated by Strong Cation Exchange (SCX) chromatography and some of the initial fractions were further used for reversed phase nano-HPLC separation and MS/MS analysis using an Orbitrap Instrument. Fractions 2-5 (2'-12'), 6-8 (12' -18') and 9-11 (18'-24'; each fraction corresponds to 2 minutes elution time) were mixed together and analyses with the Orbitrap instrument.

For the data Processing and Result Validation, the exported files were submitted for protein identification and post-translational modification characterization using Mascot 2.3 software. The databases used the K12 *E. coli* Reference Proteome subset of the UniprotKB/Swiss-Prot. The search parameters are as follow:

- The parent and fragment mass tolerance was defined at 10 ppm and 0.7Da respectively.

- Trypsin with 8 possible missed cleavages was defined for the endoproteolytic enzyme.

- Cys-carbamidomethylation was defined as fix modification.

- Met oxidation, N-term Pyro-Gln, Formyl (Protein N-term), Acetyl (Protein N-term) and D3-Acetyl (Protein N-term) were defined as variable modifications.

2.6 Microscopic methods

2.6.1 Counting of protoplasts

20 µl protoplasts solution was on the hemocytometer and counted under the light microscope. 8 big squares (each contains 16 small squares) were counted. The concentration of protoplast was calculated following formula:

Concentration: total protoplasts in 8 big squared*5000/8 = protoplasts/ml

2.6.2 CLSM Analysis

The localizations of fused-proteins in Arabidopsis protoplasts or *Nicotiana benthamiana* leaves were analyzed using Confocal LSM 510 META microscope. YFP fluorescence was excited with 514 nm, emission was recorded with a 530-560 nm band-pass filter, and RFP was excited with 543 nm and emission was recorded with a 560-615 nm band-pass filter. Chlorophyll auto-fluorescence was excited with different lasers, normally 488 nm, and emission collected within 647–745 nm. Images were further analyzed using LSM 510 software and Adobe Photoshop 7.0 (section 2.1.8).

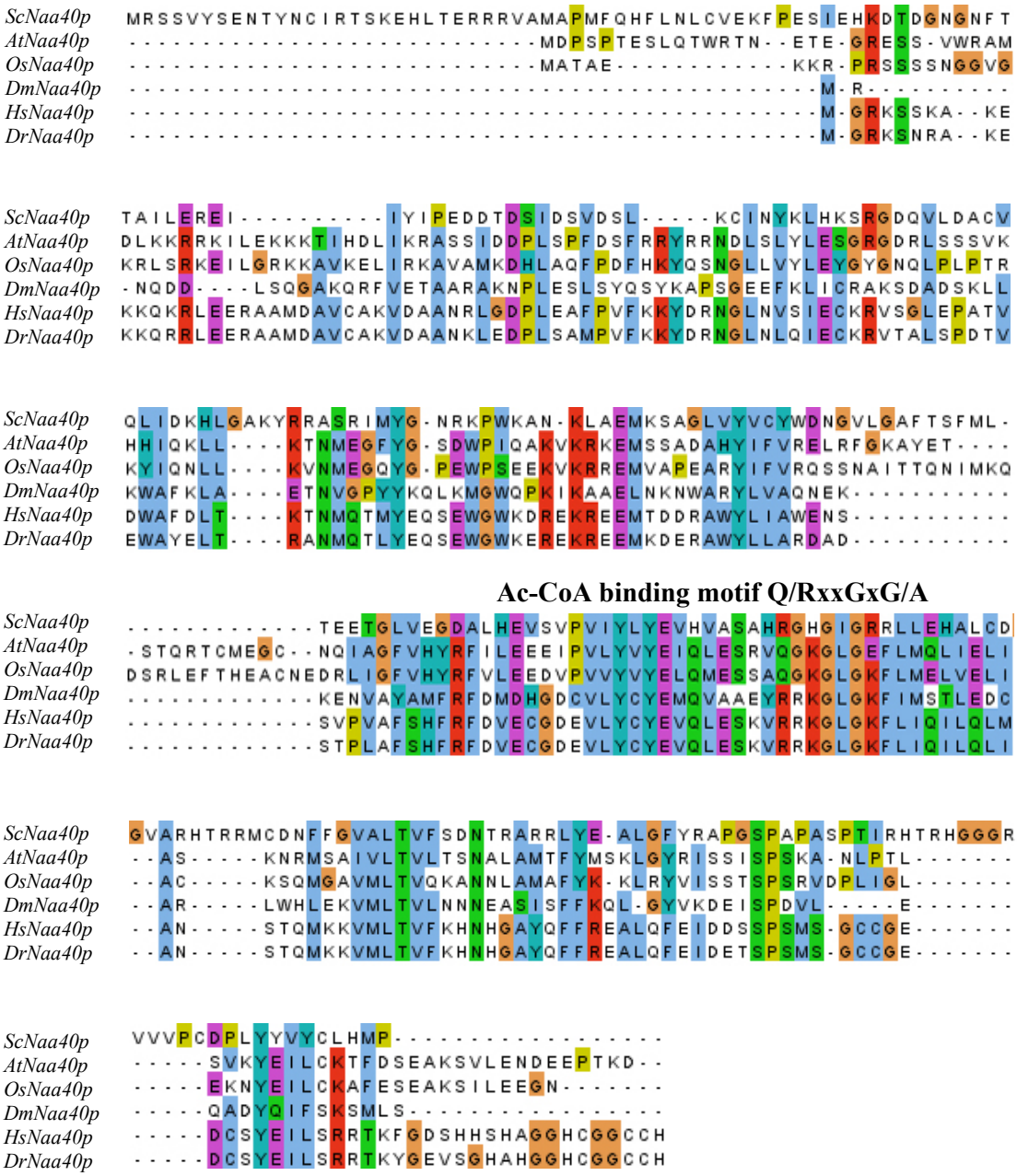
3. RESULTS

3.1 Identification and characterization of Arabidopsis NatD

3.1.1 Identification of Arabidopsis N$^{\alpha}$-acetyltransferase Naa40p by sequence similarity

Naa40p is believed to be conserved among eukaryotes with respect to protein sequence, subcellular localization and substrate specificity. To search for Arabidopsis homologue of

A



B

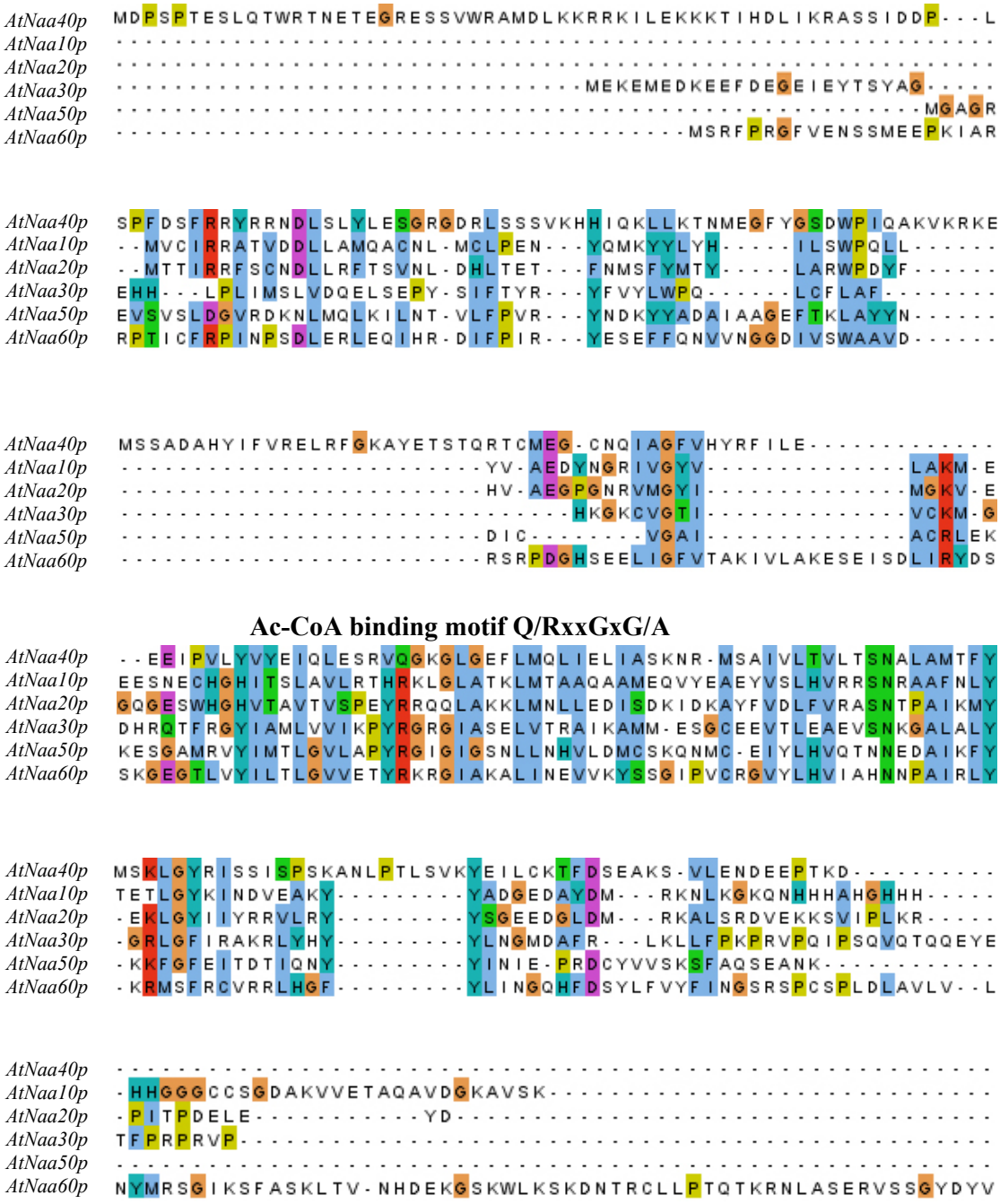


Figure 3.1 Amino acid sequence alignments of AtNaa40p and other NATs.

(A) Amino acid sequence alignment of Naa40 from *Arabidopsis thaliana* (At), *Saccharomyces cerevisiae* (Sc), *Oryza sativa* (Os), *Drosophila melanogaster* (Dm), *Danio rerio* (Dr) and *Homo sapiens* (Hs). (B) Amino acid sequence alignment of Naa40 and the catalytic subunit of other NATs of *Arabidopsis thaliana*. The conserved acetyl coenzyme A (Ac-CoA) binding motif Q/RxxGxG/A, where x represents any amino acid, is indicated in both alignments. The red line above the alignment designates for stretches of amino acid residues. The red background indicates basic residues, purple indicates acid residues, orange indicates proline, blue indicates hydrophobic residues, green indicates polar residues, and turquoise indicates histidine and tyrosine residues. The alignments were created using Clustal Omega (Goujon et al., 2010; Sievers et al., 2011), amino acids were highlighted using Jalview (Waterhouse et al., 2009).

the yeast Naa40p and human Naa40p, these two protein sequences were used as query sequences in TAIR BLAST searching against Arabidopsis proteome. In both cases, At1g18335, a protein which has a theoretical molecular mass of 31059.3 Da, was pointed out as a putative Naa40p with 28%/43% and 34%/52% identity/similarity to yeast and human Naa40p, respectively.

The N^{α}-acetyltransferase and other member of GNAT family proteins are characterized by four conserved regions (motif A-D) spanning more than 100 residues (Neuwald and Landsman, 1997). Motif A, which is the longest and the most conserved, is essential for enzyme activity. The short amino acid sequence Q/RxxGxG/A of motif A involves in Ac-CoA binding. Point mutations in any of these residues lead to more than 90% reduced activity of enzymes (Tercero and Wickner, 1992; Coleman et al., 1996; Hilfiker et al., 1997; Dutnall et al., 1998; Smith et al., 1998). As shown in the Fig. 3.1A, the conserved Ac-CoA binding motif is present in plant Naa40p (Arabidopsis and rice) strongly indicating a potential of N^{α}-acetyltransferase activity. The conservation of this conserved motif and an overall sequence similarity among yeast, Arabidopsis and human point out At1g18335 as the orthologue of yNaa40p.

Besides the conservation, all NATs in general and Naa40p in particular display unique characteristics of its own. AtNaa40p has unique stretches which do not present in other Arabidopsis NATs (Fig. 3.1B). The first glutamine (Q) amino acid residue of the conserved Ac-CoA binding motif is typical for plant kingdom, and some stretching sequences are specific for yeast (Fig. 3.1A).

3.1.2 Over-expression and purification of recombinant AtNaa40p

In order to express and purify AtNaa40p for N^{α}-acetyltransferase activity assay, *ATNAA40* cDNA was amplified using PCR with primers number 2392 and 2393, the restriction sites *NcoI* and *KpnI* were added into primers for cloning into pETM41 expression vector (Table S2). In pETM41 expression system, AtNaa40p was fused with maltose binding protein (MBP) and 6 histidine residues (His-Tag) at the N-terminus (Fig. S4). MBP was able to increase the solubility of recombinant AtNaa40p and His-Tag is necessary for purification using HiTrap Chelating HP column. The protein was expressed in *E. coli* Rosetta, and purified from the protein extract by immobilized metal affinity chromatography (IMAC) (section 2.3.5). The SDS-PAGE of different fractions of protein purification was shown in the Fig. 3.2. The increasing amount of MBP-AtNaa40p displayed the time dependent

manner indicating the successful expression of this protein in *E. co*li (Fig. 3.2A). A very small amount of MBP-AtNaa40p was found in the flow through fraction indicating that most of protein of interest was bound to the column. A large amount of unspecific proteins without His-Tag were washed away during washing fractions 1-5. MBP-AtNaa40p was collected in the time of elution fraction 2, 3 and 4 with concentrations 0.57 µg/µl, 1.72 µg/µl and 0.75 µg/µl, respectively (Fig. 3.2B). Since MBP-AtNaa40p from the elution fraction 4 is more purified and the concentration is high enough, the protein extract from this fraction was used for N^{α}-acetyltransferase activity assay.

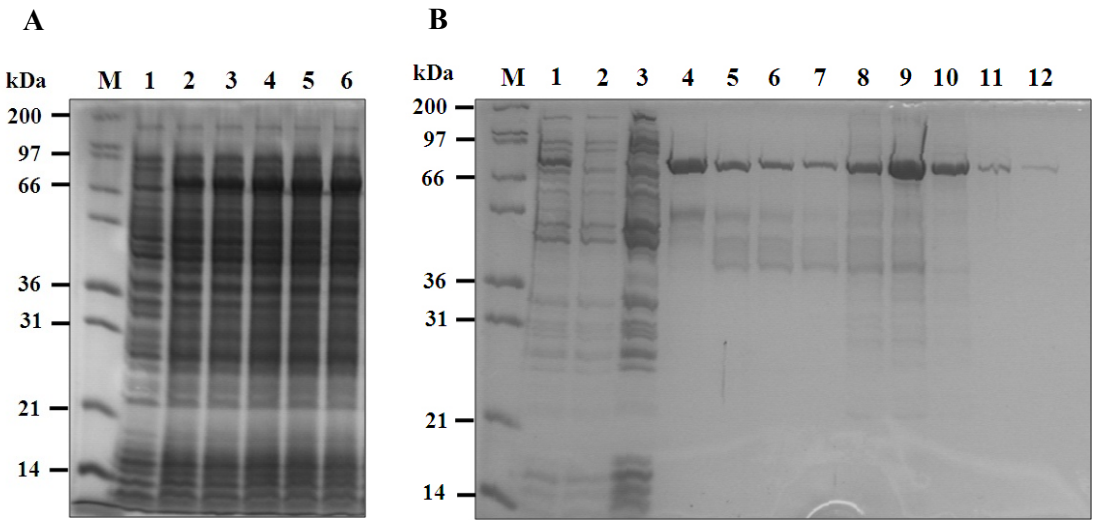


Figure 3.2 Expression and purification of recombinant AtNaa40p.

Proteins from the different fractions of purification procedure were separated using SDS-PAGE and visualized by Coomassie staning (section 2.5.8). (A) Expression of MBP-AtNaa40p over time after induction. 10 µl of protein extract from 1ml cell culture (section 2.3.9) from different time points of induction was loaded. (M) Mark12 protein ladder, (1) no induction, (2-6) 1-5 hours induction. (B) Protein purification from 300 ml overnight cell culture. (M) Mark12 protein ladder, (1) Crude extract – 10 µg, (2) Flow through extract – 10 µg, (3) Wash fraction 1 – 10 µg, (4-6) Wash fractions 5,10,15 – maximum 2 µg of protein was loaded, (7-12) Elution fractions 1-6 – maximum 2 µg of protein was loaded.

3.1.3 Arabidopsis Naa40p displays N^{α}-acetyltransferase activity

Although the conservation of amino acid sequence clearly supports AtNaa40p as a novel Arabidopsis NAT of the NatD type, its N^{α}-acetyltransferase activity has to be proven. Human NatD and yeast NatD were reported to N-terminally acetylate proteins histone H4 and histone H2A (Song et al., 2003; Polevoda et al., 2009b; Hole et al., 2011). In order to assess whether AtNaa40p is a true NAT, oligopeptides were designed for *in vitro* N^{α}-

acetyltransferase activity assay. All custom-synthesis peptides contain seven unique amino acids at the N-terminus, since these residues are the major determinants for substrate specificity of N^{α}-acetyltransferase. The next 17 amino acids are identical to the ACTH peptide sequence to maintain a positive charge facilitating peptide solubility and effective isolation by cation exchange Sepharose beads. The ACTH derived lysines were replaced by arginines to eliminate any potential interferences of N^{ε}-acetylation (Starheim et al., 2009).

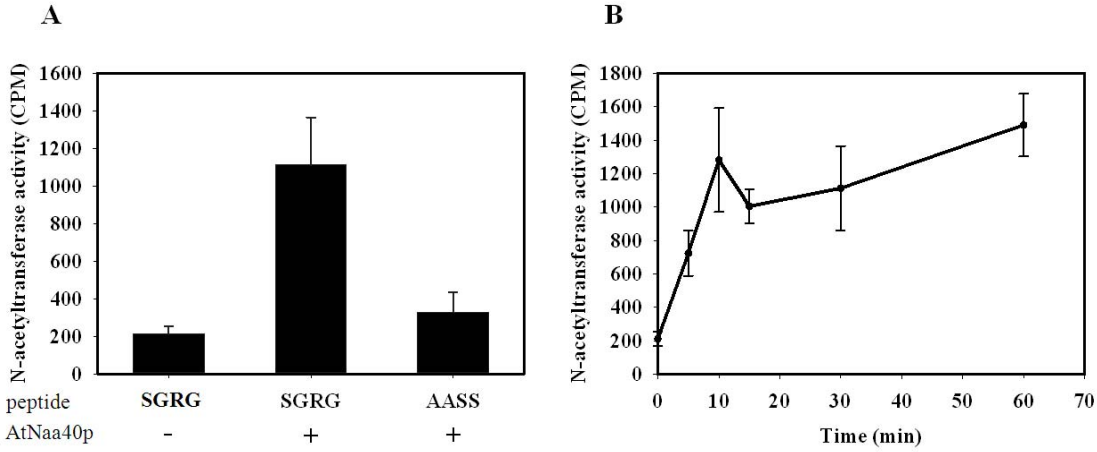


Figure 3.3 AtNaa40p displays N^{α}-acetyltransferase activity.

0.2 µg of purified BMP-AtNaa40p was incubated with 2 µl of 2.47 MBq/ml ^{3}H-Ac-CoA, 10 µl of 400 µM selected peptides. Peptides represent for different protein N-termini and the first four N-terminal residues are shown for each substrate. The acetyl incorporation was measured by scintillation counter. Error bars indicate standard error for two independent experiments. (A) AtNaa40p displays sequence-specific N^{α}-acetyltransferase activity. (B) AtNaa40p shows time course enzymatic kinetics.

The N^{α}-acetyltransferase activity assay proved AtNaa40p is a genuine NAT. The purified recombinant AtNaa40p was able to acetylate a 24 amino acid containing peptide in a time dependent manner (Fig 3.3B). The ability to acetylate target peptide was dependent on sequence requirements within the first seven amino acids. The histone H4 mimicked peptide was N-terminally acetylated as expected, whereas a typical NatA type substrate AASS-peptide was not acetylated (Fig. 3.3A).

In addition, Hole et al. (2011) reported that hNaa40p is also able to acetylate peptides starting with glycine (G) and methionine (M). In agreement with this study, in another experiment, ARK1 peptide (MVNA) was acetylated as the same level as histone H4 mimicked peptide (SGRG) by AtNaa40p (Fig. 3.4). All together, the data prove that

AtNaa40p is a genuine N^{α}-acetyltransferase and its function seems to be conserved from yeast and human with respect to Nt-acetylation of histone H4.

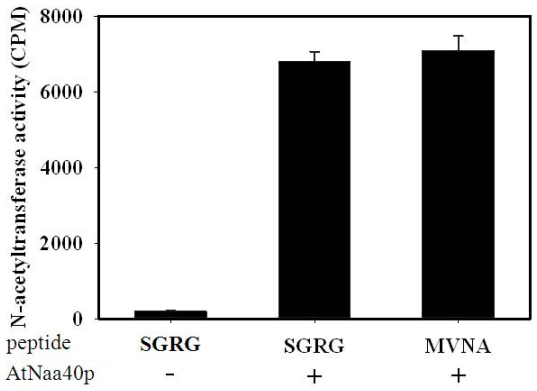


Figure 3.4 AtNaa40p N-terminally acetylates peptides with SGRG and MVNA at the N-terminus.
0.4 µg of BMP-AtNaa40p was incubated with 2 µl of 2.47 MBq/ml ^{3}H-Ac-CoA, 10 µl of 400 µM selected peptides. Peptides represent for different protein N-termini and the first four N-terminal residues are shown for each substrate. The acetyl incorporation was measured by scintillation counter. Error bars indicate standard error for two independent experiments.

3.1.4 Subcellular localization of AtNaa40p

The subcellular localization of AtNaa40p homologues from yeast and human has already been studied. yNaa40p localizes to the cytoplasm and hNaa40p is located both in cytoplasm and in nucleus (Polevoda et al., 2009b; Hole et al., 2011). *In silico* analysis predicted that AtNaa40p contains a putative nuclear localization signal (NLS) along with conserved acetyltransferase domain (Fig. 3.5). According to the cNLS mapper (Kosugi et al., 2009), the putative NLS of AtNaa40p (DLKKRRKIL) located between amino acids 29-38 with cut-off score of 5 indicating the localization both to the nucleus and to the cytoplasm. In addition, the Arabidopsis Subcellular Database (SUBA) (Heazlewood et al., 2007) predicted the localization of AtNaa40p only to the nucleus (Table 3.1).

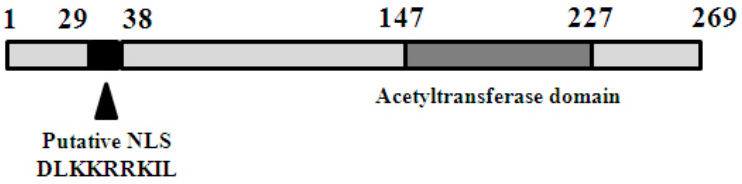


Figure 3.5 Schematic model of AtNaa40p.

AtNaa40p is a protein of 269 amino acids with a putative NLS between amino acids 29-38. The acetyltransferase domain locates between amino acids 147-227. The putative NLS was predicted by cNLS mapper (Kosugi et al., 2009). SMART was used to analyze the acetyltransferase domain (Schultz et al., 1998; Letunic et al., 2012).

Table 3.1 SUBA prediction of AtNaa40p subcellular localization

Predictors	Subcellular localization
iPSORT	no data
LOCtree	no data
MitoPred	no data
MultiLoc	no data
PeroxP	no data
Predotar	no data
SubLoc	**nucleus**
Mitoprot 2	no data
TargetP	no data
WoLFPSORT	**nucleus**

In order to address experimentally the subcellulat localization of AtNaa40p, EYPF was fused to the C-terminus of AtNaa40p and transiently expressed in Arabidopsis protoplasts. *ATNAA40* cDNA was amplified by PCR using primers number 2332 and 2196, the restriction sites *KpnI* and *XbaI* were added into primers for cloning into pFF19-EYFP expression vector (Table S2). The pFF19-EYFP-AtNaa40 and pURT2-Ubi-RFP (RFP-cytosolic marker) (Table S2) were co-transfected into prepared Arabidopsis protoplasts for co-localization (sections 2.2.6 and 2.2.7).

The localization of AtNaa40-EYFP fused protein and RFP-cytosolic marker was shown in the (Fig. 3.6). A part of AtNaa40-EYFP was present in the cytoplasm but a larger amount was detected in the nucleus (Fig. 3.6A). The cytosolic marker equally localized to the nucleus and to the cytoplasm (Fig. 3.6B). The perfect overlay of YFP signal and RFP marker (Fig. 3.6C) clearly indicates the localization of AtNaa40p to the cytoplasm.

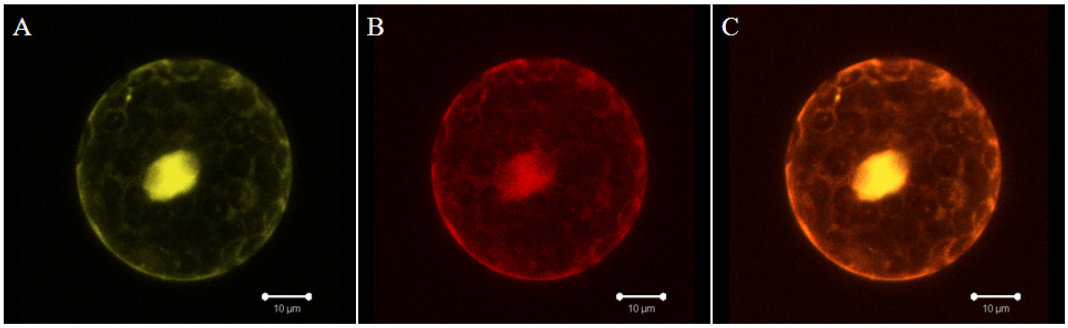


Figure 3.6 Subcellular localization of AtNaa40-EYFP fused protein in Arabidopsis protoplasts.

(A) AtNaa40-EYFP fused protein was detected both in the cytoplasm and in the nucleus. (B) Signals of cytosolic-RFP marker presented in the nucleus and in the cytoplasm. (C) Overlay of YFP and RFP channels.

The cell nucleus is separated to cytoplasm by nuclear envelope which is penetrated with nuclear pore complexes (NPC). Each NPC contains one or more open aqueous channels which control the transport of molecular between the nucleus and the cytosol. Water-soluble molecules that are smaller than a certain size can passively diffuse through the channels. Small molecules (5000 daltons or less) diffuse in so fast that the nuclear envelope can be considered to be freely permeable to them. A protein of 17.000 daltons takes 2 minutes to equilibrate between the cytosol and nucleus, while a protein of 44.000 daltons takes 30 minutes. A globular protein larger than about 60,000 daltons seems hardly able to enter the nucleus at all. To enter the nuclear, larger proteins need to be imported by NPC by using energy and requiring a nuclear import signal (NLS). The RFP cytosolic marker is a 27.000 daltonls protein which can easily enter to nucleus. That is the reason why the RFP signal of cytosolic marker was detected in the nucleus. However, AtNaa40-EYPF fused protein with a molecular mass approximate 60.000 daltons is too big to pass the NLS by passive diffusion. To enter the nucleus, it needed an active transport with the participation of nucleoporins and required a NLS and ATP. As mention above, AtNaa40p contains a putative NLS at the N-terminus which allows it to access to the nucleus. *In silico* analysis and subcellular localization studies demonstrate that AtNaa40p not only localizes to the cytoplasm but also to the nucleus.

3.1.5 Identification and characterization of *atnaa40* mutants

3.1.5.1 Identification of *atnaa40* mutants

To further characterize the significance of NatD in *Arabidopsis thaliana*, two independent mutant lines with T-DNA insertion at locus *AT1G18335* were isolated from the SALK collection. The first mutant (*atnaa40-1*, SALK_152065C) contains T-DNA insertion at the first exon, and the T-DNA insertion of the second mutant (*atnaa40-2*, SALK_081588) locates at the first intron (Fig. 3.7). The homozygote plants were identified by PCR genotyping with gene specific primers and T-DNA insertion primers (section 2.2.6). All primers were designed using T-DNA Primer Design online software (http://signal.salk.edu/tdnaprimers.2.html). Total genomic DNA extraction from Arabidopsis and PCR program were performed as described in (section 2.4.1 and 2.4.2). The homozygote mutants with two T-DNA insertions in both alleles had a T-DNA specific band, whereas the wild type plant with no T-DNA insertion had a gene specific band (Fig. 3.8). The homozygote plants were used for further experiments.

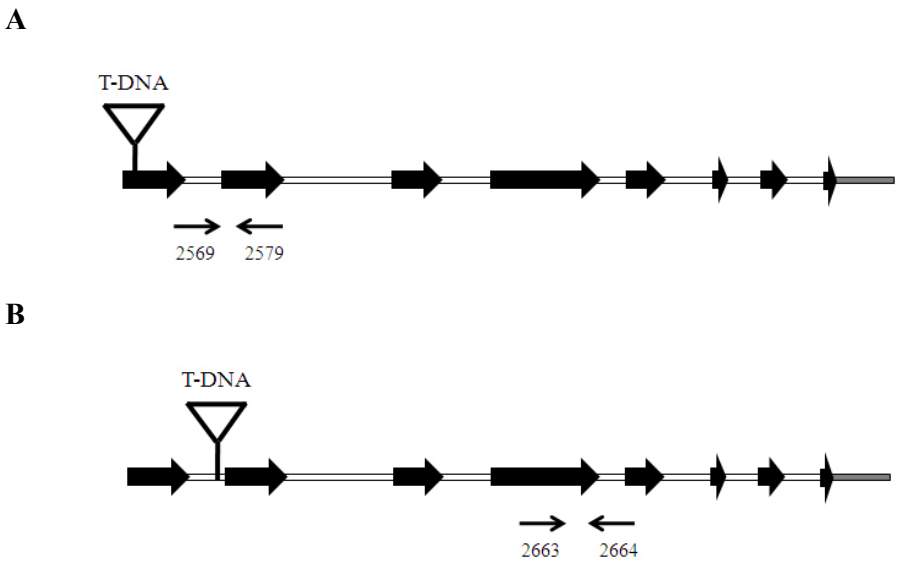


Figure 3.7 Structure of *ATNAA40* with T-DNA insertions.

(A) *atnaa40-1* with T-DNA insertion at the first exon. (B) *atnaa40-2* with T-DNA insertion at the first intron. Big black arrow indicates exon, white bar indicates intron, grey bar indicates un-translated region (UTR) and small back arrow indicates qRT-PCR primers.

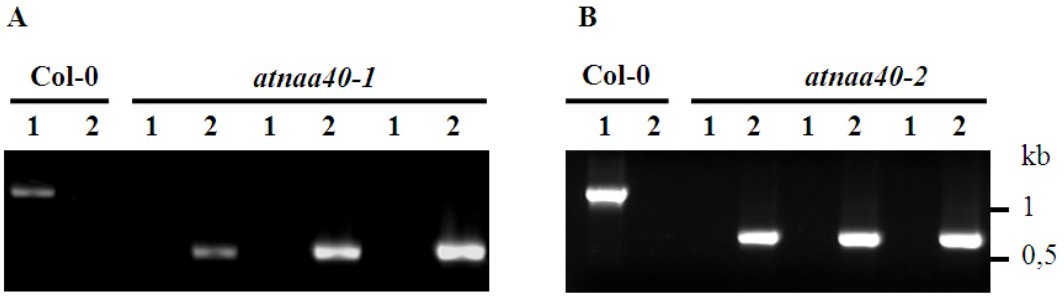


Figure 3.8 Genotyping of *atnaa40* T-DNA insertion lines.

PCR genotyping of *atnaa40* mutant lines was run with two pairs of primers, gene specific primers (1) and T-DNA insertion specific primers (2). (A) Genotyping of *atnaa40-1* mutant: (1) PCR with primer number 2499 and 2500, (2) PCR with primer number 2499 and 1401. (B) Genotyping of *atnaa40-2* mutant: (1) PCR with primer number 2425 and 2426, (2) PCR with primer number 2325 and 1401.

3.1.5.2 The transcription level of *ATNAA40* in *atnaa40-1* and *atnaa40-2* are altered

To evaluate the effect of T-DNA insertion on the function of *ATNAA40* gene, qRT-PCR with different primers (Fig 3.7) was performed (section 2.4.8). The T-DNA insertion at the first exon impelled the expression of *ATNAA40* almost two fold than normal (Fig. 3.8A). In contrast, the expression of *ATNAA40* was severely disrupted by the insertion of T-DNA at the first intron (only about 2.5 % remaining) (Fig. 3.8B).

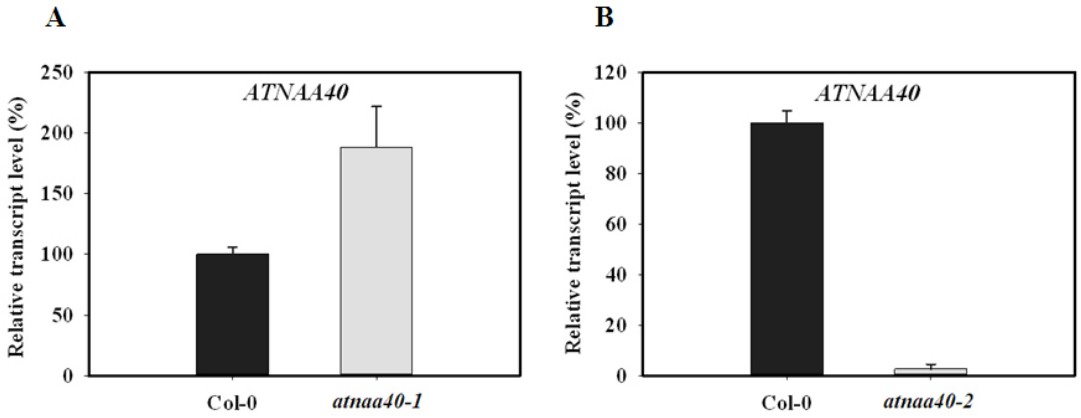


Figure 3.9 Transcription level of *ATNAA40* in the T-DNA insertion lines.

(A) In the *atnaa40-1* mutant, the transcription of *ATNAA40* was 2 fold higher than the wild type. qRT-PCR was run with primers 2569 and 2579. (B) T-DNA insertion at the first intron severely disrupted the expression of *ATNAA40* in *atnaa40-2* mutant. qRT-PCR was run with primers 2663 and 2664. Error bar indicates standard error (n=3).

3.1.5.3 Phenotypes of *atnaa40-1* and *naa40-2* mutants

NatD (Naa40p) has been identified and characterized both in yeast and in human. So far yNaa40p and hNaa40p are known to be responsible for the N-terminal acetylation of histones H2A and H4. The *ynaa40*-deletion strain did not show any phenotype under the normal growth conditions (Song et al., 2003). However, in the media containing NaCl or an inhibitor of transcription (3-amino-1, 2, 4-triazole) this deletion strain induces some minor growth defects (Polevoda et al., 2009b).

Similar to yeast, two T-DNA insertion mutants (*atnaa40-1* and *atnaa40-2*) did not show any observable phenotypes in comparison to the wild type under the normal growth conditions (section 2.2.2). Regardless whether the *ATNAA40* coding product in the *atnaa40-1* mutant is properly functioning or not, the over-expression and knockdown of *ATNAA40* both resulted in normal growth, normal development and ability of producing seeds (Fig. 3.10).

To further investigate the phenotypes of *atnaa40* mutants, seeds from *atnaa40-1*, *atnaa40-2* and wild type plants were germinated *in vitro* on the MS media or MS media supplemented with different concentrations of NaCl. On the MS media, both mutants and the wild type seeds germinated at the same rate. However, on the MS media supplemented with NaCl, *atnaa40-1* germinated less efficiently than the others (Fig. 3.11). Interestingly, *atnaa40-1* is the mutant in which the mRNA steady state level of *ATNAA40* is 2 fold higher than in the wild type. There are some possibilities that could explain for this unexpected result. Firstly, T-DNA insertion at the first exon resulted in the non-functional protein that does not have enzymatic activity and 2.5% expression of *ATNAA40* in the *atnaa40-2* mutant was enough to fulfill the function of AtNatD at the certain time. Secondly, there are other insertions of T-DNA in the *atnaa40-1* genome which have effects on the phenotypes. Less likely, the over-expression of *ATNAA40* could have the negative effects on the phenotypes of *atnaa40-1* mutant.

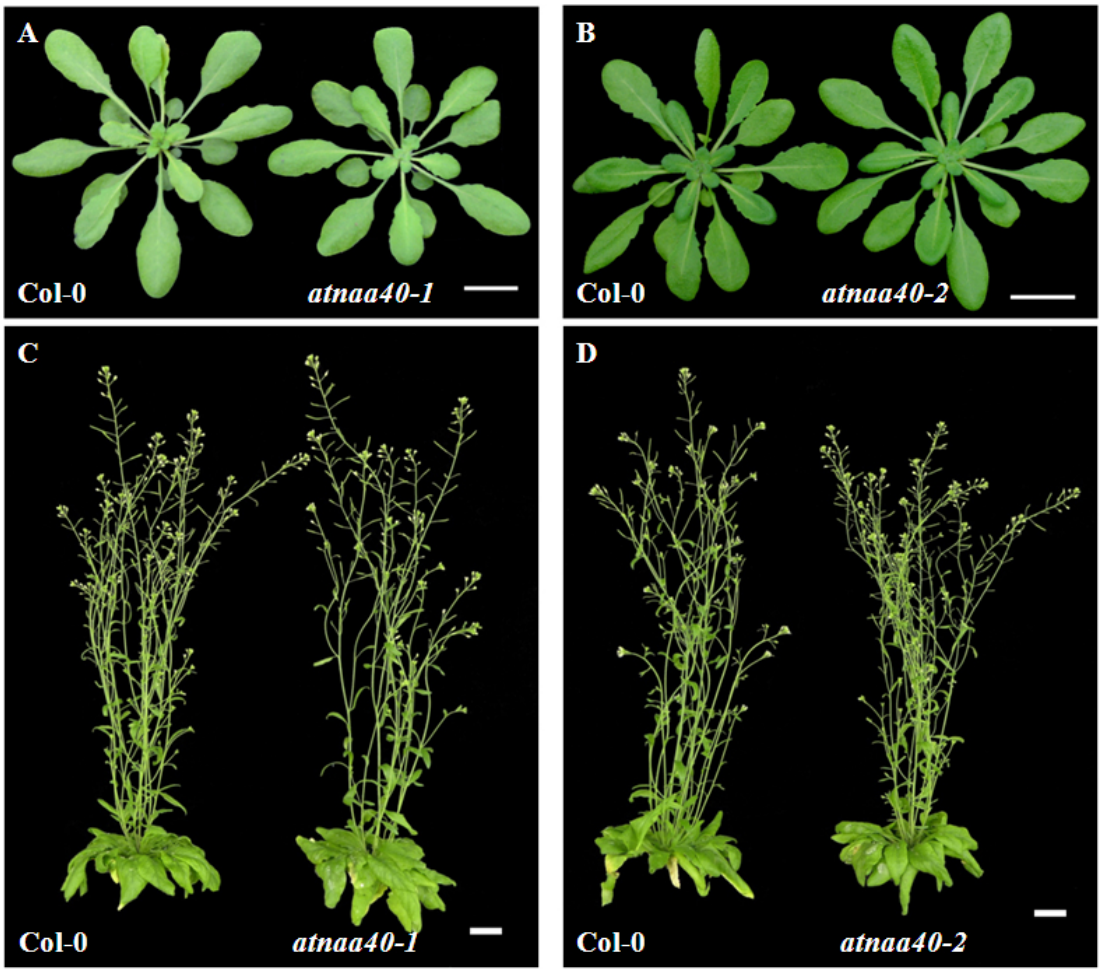


Figure 3.10 Phenotypes of *atnaa40* mutants under normal conditions.

Seeds were planted in the soil under short day conditions, after 10 weeks, plants were transferred to the long day chamber for producing seeds. (A and B) Seven week-old plants were under short day conditions. (C and D) Twelve week-old plants were flowering and producing seeds in the long day chamber. White bar indicates scale of 2cm.

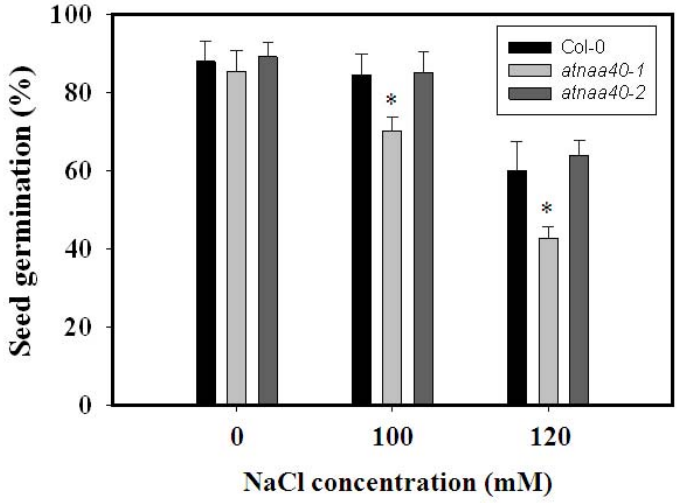


Figure 3.11 Germination of *atnaa40* mutant seeds *in vitro*.

Seeds from *atnaa40-1, naa40-2* and the wild type were sterilized and germinated on the MS media or MS media supplemented with 100 mM or 120 mM NaCl. Two weeks after plating, the seed germination was counted. Error bar indicates standard error (n=3 or 5). Star indicates the significant difference ($p < 0.05$).

3.1.5.4 Phenotypes of *atnaa40* mutants under UV-c stress

Histones H2A and H4, two substrates of Naa40p, were reported to play an important role in DNA double-strand break repair (Bilsland and Downs, 2005; Bonisch and Hake, 2012). The intact histone N-tail is required for normal function of the cell and N-terminal serine acetylation alters the charge of histone N-tail (Johnson et al., 1990; Kayne et al., 1988; Polevoda et al., 2009b). It would be possible that the acetylation of N-termini of histones H2A and H4 might contribute a part in DNA damage repair. In y*naa40* mutant, histones H2A and H4 are not N-terminally acetylated (Song et al., 2003).

In order to investigate the sensitivity of *atnaa40* mutants to DNA damaging agent, three week-old plants were exposed to UV-c for different periods of time (section 2.2.5). However, there were no observed differences among the wild type and the mutants. All plants which were exposed longer than 20 minutes were faded and gradually died after radiation (Fig. 3.12 B-E). Only 20 minutes radiated plants were still viable although old leaves died. These plants continuously grew and looked normally after two weeks recovery (Fig. 3.12 A and Fig. 3.12 F).

Figure 3.12 Phenotypes of *atnaa40* mutants under UV-c tress.

Three week-old plants were exposed to UV-c in different periods of time. After radiation, plants were recovered in the short day conditions. Pictures were taken 5 days after recovery. (A) 20 minutes UV-c exposed plants. (B) 40 minutes UV-c exposed plants. (C). 60 minutes UV-c exposed plants. (D) 90 minutes UV-c exposed plants. (E) 120 minutes UV-c exposed plants. (F) 20 minutes UV-c exposed plants were growing after two weeks recovery. White bar indicates scale of 2cm.

3.1.5.6 Status of protein N-terminal acetylation in the *atnaa40* mutants

Up to now, NatD was reported to acetylate only protein histones H4 and H2A (Song et al., 2003; Polevoda et al., 2009b). However, the lack of identified substrates does not exclude their existence. In order to investigate the potential acetylation of other proteins by AtNaa40p, the free N-terminus protein status in *atnaa40* mutants was analyzed. It was reported that the unacetylated protein is able to react with NBD-Cl to provide a stable fluorescence, whereas the acetylated form is not (Bernal-Perez et al., 2012). Total protein was extracted from leaves of six week-old Arabidopsis plants (section 2.5.1). To avoid any interference in reaction, the free amino acids and other reagents were eliminated by using PD SpinTrapTM G-25 column (section 2.5.2). The purified protein extract was used to perform the NBD-Cl fluorescence assay (section 2.5.6).

The fluorescence intensities of overnight reaction solutions were shown in Fig. 3.13. The wild type and both *atnaa40* mutants had similar fluorescent values indicating the same level of overall protein acetylation. This result supports to a certain degree the theory of Naa40p acetylating only protein histones H4 and H2A.

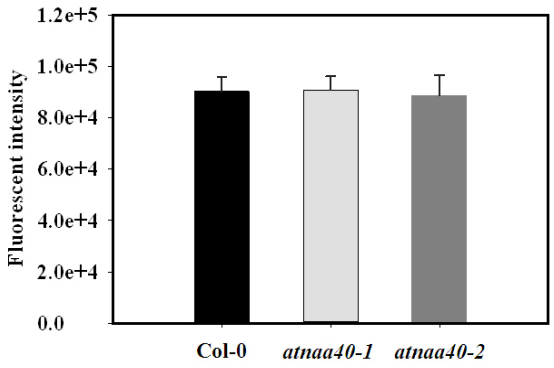


Figure 3.13 Status of protein N$^{\alpha}$-acetylation in the *atnaa40* mutants.

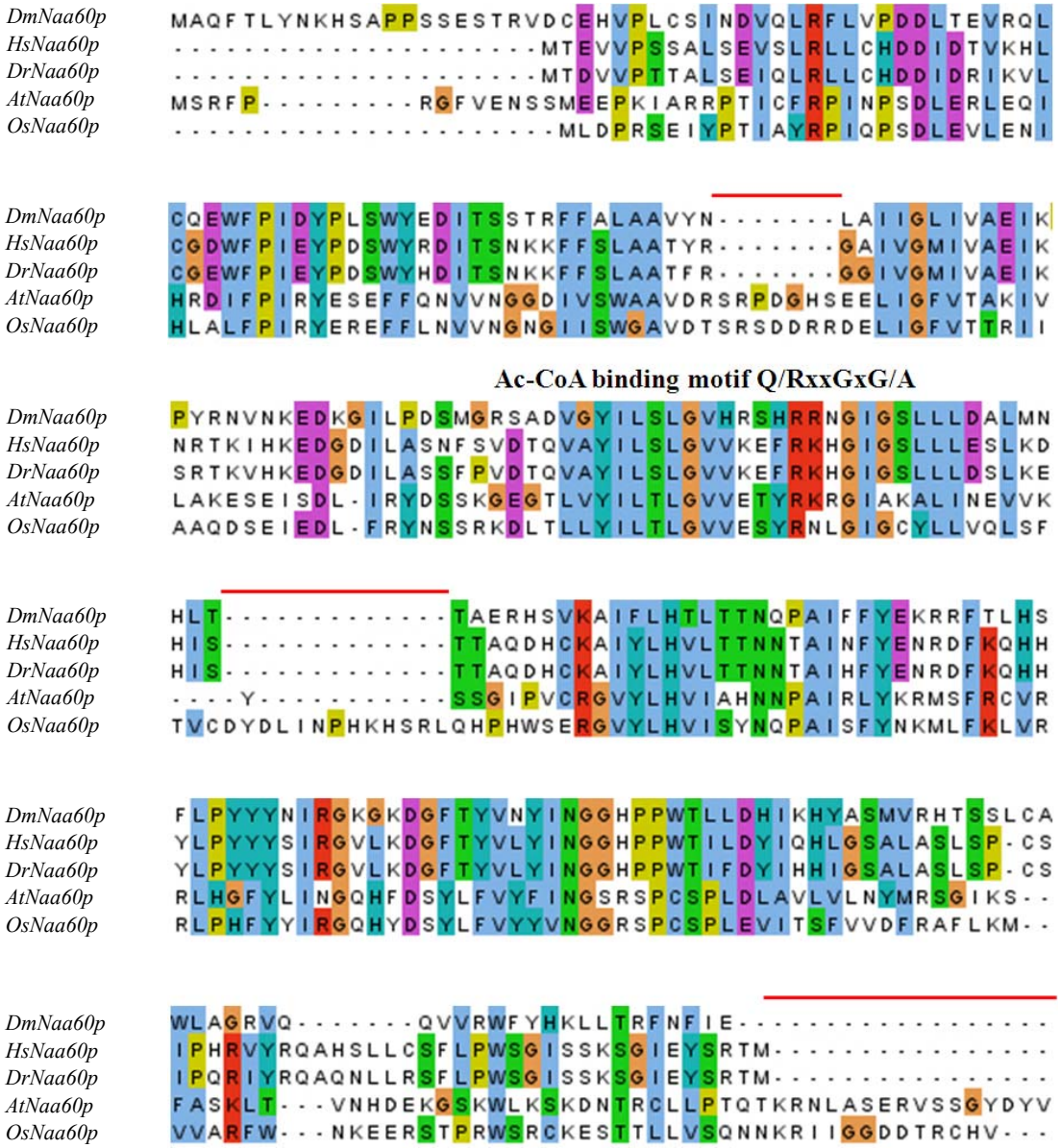
7,6 µg of purified protein extract was reacted with 0.5 mM NBD–Cl in 50 mM sodium citrate buffer overnight at RT. Overnight reaction solution was exited at 470 nm and fluorescence intensity was measured at 520 nm. Error bar indicates standard error (n=4).

3.2 Identification and characterization of Arabidopsis NatF

3.2.1 Identification of Arabidopsis N$^{\alpha}$-acetyltransferase Naa60p by sequence similarity

Human NatF (hNatF/hNaa60p) has been recently identified as a novel type of NAT which is unique for higher eukaryotes and displays particular substrate specificity (Van Damme et al. 2011c). Since the protein sequence of Naa60p is conserved in higher eukaryotes, in a search for homologue of Naa60p in plants the amino acid sequence of hNaa60p was used as a query searching against Arabidopsis proteome. As a result of blasting, At5g16800 was

A



B

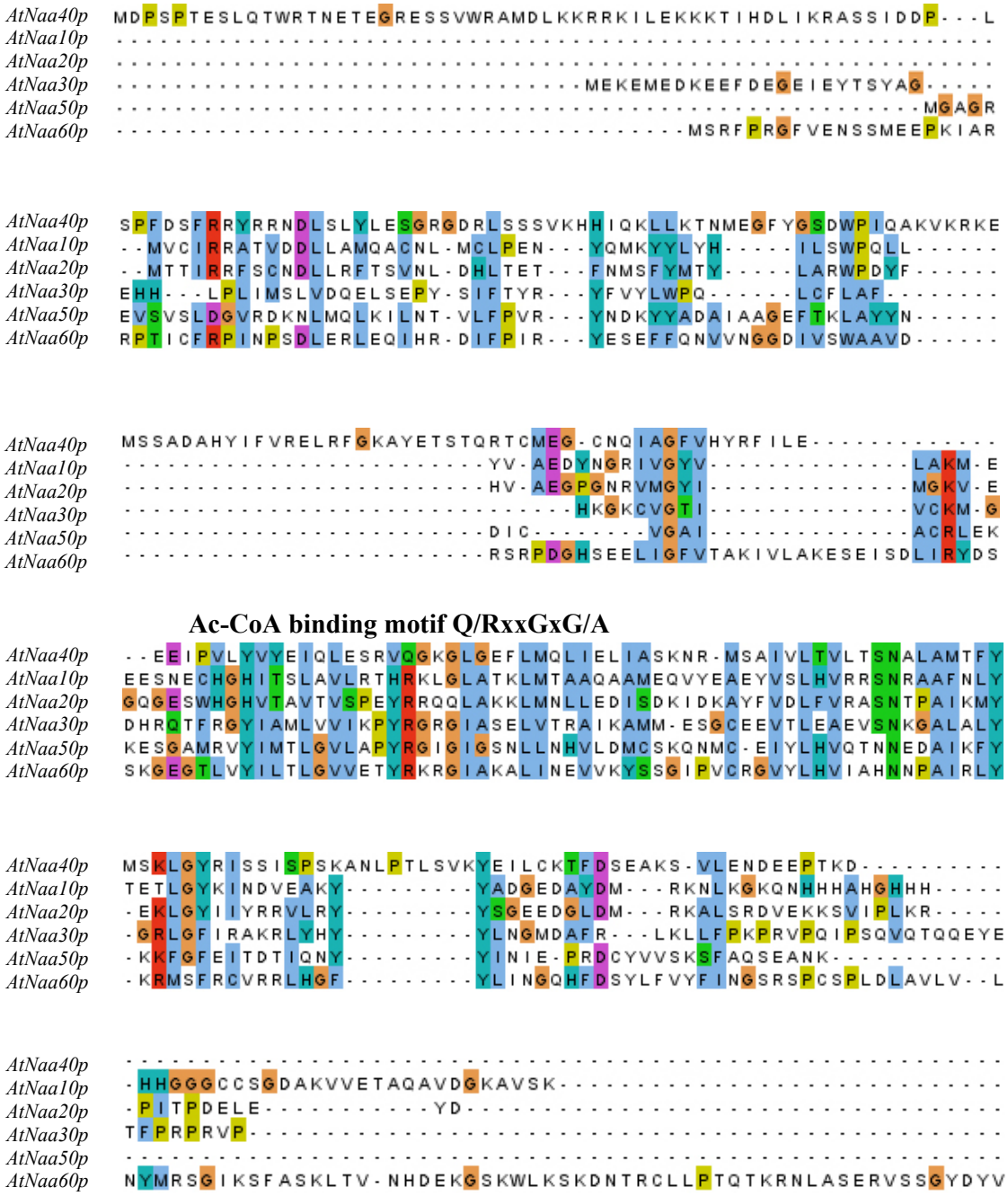


Figure 3.14 Amino acid sequence alignments of AtNaa60p and other NATs.

(A) Amino acid sequence alignment of Naa60 from *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Drosophila melanogaster* (Dm), *Danio rerio* (Dr) and *Homo sapiens* (Hs). (B) Amino acid sequence alignment of Naa60 and the catalytic subunit of other NATs of *Arabidopsis thaliana*. At5g16800.2 was used as AtNaa60 in alignments. The conserved Ac-CoA binding motif Q/RxxGxG/A, where x represents any amino acid, is indicated in both alignments. The red line above the alignment designates for stretches of amino acid residues. The red background indicates basic residues, purple indicates acid residues, orange indicates proline, blue indicates hydrophobic residues, green indicates polar residues, and turquoise indicates histidine and tyrosine residues. The alignments were created using Clustal Omega (Goujon et al., 20010; Sievers et al., 2011), amino acids were highlighted using Jalview (; Waterhouse et al., 2009).

pointed out as a putative NatF in Arabidopsis. According to TAIR, At5g16800 locus has three different gene models *AT5g16800.1, AT5G16800.2* and *AT5G16800.3*. Since the isoform At5g16800.2 has the highest identity and similarity (28% and 48%) to hNaa60p, it was considered as AtNaa60p and *AT5G16800.2* DNA sequence was used for further experiments. The alignment of predicted Naa60p form different species was shown in Fig. 3.14A. The Ac-CoA binding motif (Q/RxxGxG/A) which is essential for the function of N^{α}-acetyltransferase (Tercero and Wickner, 1992; Coleman et al., 1996; Hilfiker et al., 1997; Dutnall et al., 1998; Smith et al., 1998) is conserved in all organisms. The presence of Ac-CoA binding motif and amino acid sequence conservation strongly indicates At5g16800 as a potential N^{α}-acetyltransferase.

In addition, Naa60p from different species also display the unique characteristic. Some sequence stretches are specific for plant Naa60p (Fig. 3.14A). Within plant NATs, AtNaa60p shows the differences to others with a stretch and an additional sequence at the C-tail (Fig. 3.14B).

3.2.2 Over-expression and purification of recombinant AtNaa60p

Since AtNaa40p was successfully expressed and purified using bacterial expression system, the same procedure was applied for AtNaa60p. *ATNAA60* cDNA was amplified using PCR with primers number 2367 and 2339, the restriction sites *NcoI* and *BamHI* were added into primers for cloning into pETM 41 vector (Table S2 and Fig. S4). The recombinant protein was expressed in *E.coli* Rosetta and purified from the protein extract by immobilized metal affinity chromatography (IMAC) (section 2.3.5). The SDS-PAGE of different fractions of protein purification was shown in the Fig. 3.15. The increasing amount of MBP-AtNaa60p displayed the time dependent manner indicating the successful expression of this protein in *E. coli* (Fig 3.15A). During the purification, a large amount of the unspecific proteins were washed away from the column in the time between washing fraction 1 and 5. Most of MBP-AtNaa60 was collected at the elution fraction 2, 3 and 4 with concentration 0.97 µg/µl, 1.0 µg/µl and 0.48 µg/µl, respectively (Fig. 3.15B). Since AtNaa60p collected from the elution fraction 4 was more pure and the concentration was high enough, the protein extract from this fraction was used for N^{α}-acetyltransferase activity assay.

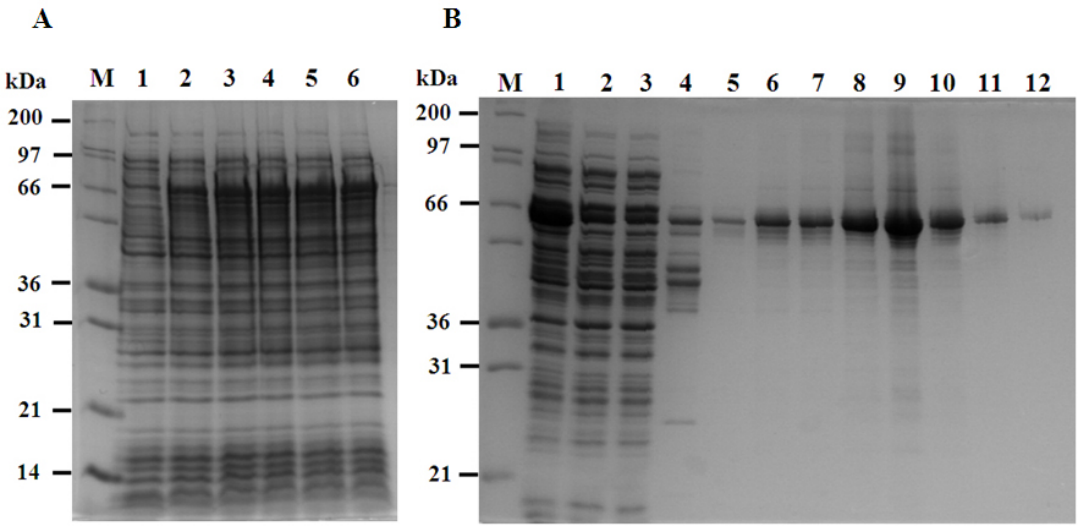


Figure 3.15 Expression and purification of recombinant AtNaa60p.
Proteins from the different fractions of purification procedure were separated using SDS-PAGE and visualized by Coomassie staning (section 2.5.8). (A) Expression of MBP-AtNaa40p over time after induction. 10 µl of protein extract from 1ml cell culture (section 2.3.9) from different time points of induction was loaded. (M) Mark12 protein ladder, (1) no induction, (2-6) 1-5 hours induction. (B) Protein purification from 300 ml overnight cell culture. (M) Mark12 protein ladder, (1) Crude extract – 10 µg, (2) Flow through extract – 10 µg, (3) Wash fraction 1 – 10 µg, (4-6) Wash fractions 5,10,15 – maximum 2 µg of protein was loaded, (7-12) Elution fractions 1-6 – maximum 2 µg of protein was loaded.

3.2.3 Arabidopsis Naa60p displays N^{α}-acetyltransferase activity

NatF is the most recently identified NAT which is representative of higher eukaryotes and shows unique substrate specificity. hNatF was reported to N-terminally acetylate methionine followed by different amino acid residues (Van Damme et al., 2011c). In order to prove AtNaa60p as a true N^{α}-acetyltransferase, purified MBP-AtNaa60p was incubated with different peptides. Each peptide was designed with variant N-termini representing for different proteins (section 3.1.3). As it was expected, AtNaa60p was able to acetylate MVNA- peptide. MVNA- is N-terminus of palmitoyltransferase AKR1 protein which is N-terminally acetylated by hNaa60p (Van Damme et al. 2011c). AASS-, which is a typical NatA type substrate was not acetylated (Fig. 3.16). Surprisingly, AtNaa60p acetylated not only MVNA- but also SGRG- (N-terminus of histone H4). N-terminal acetylaion of peptide starting with serine by AtNaa60p was unexpected since hNaa60p can only acetylate methionine residue and Naa60p was believed to be conserved among higher

eukaryotes with respect to both amino acid sequence and substrate specificity (Van Damme et al. 2011c).

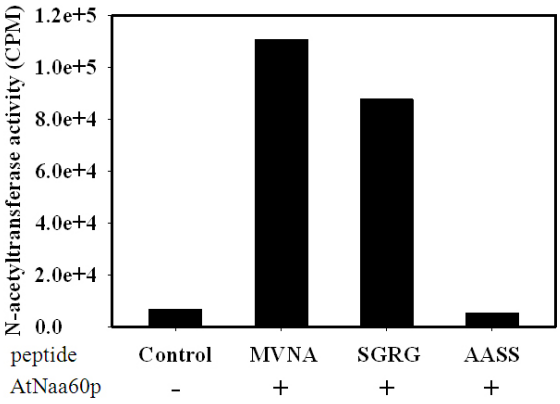


Figure 3.16 AtNaa60p displays N$^{\alpha}$-acetyltransferase activity *in vitro*.

0.5 µg of purified BMP-AtNaa60p was incubated with 2 µl of 3.7 MBq/ml ^{3}H-Ac-CoA, 10 µl of 400 µM selected peptides. Peptides represent for different protein N-termini and the first four N-terminal residues are shown for each substrate. The acetyl incorporation was measured by scintillation counter.

3.2.4 Subcellular localization of AtNaa60p

Despite the fact that Naa60p has been well characterized in human, the subcellular localization was not mentioned. *In silico* analysis revealed different possibilities of AtNaa60p subcellular localization. PredictProtein (Rost et al., 2004) predicted two putative transmembrane helices which locate in the middle and at the N-terminus of AtNaa60p (Fig. 3.17) suggesting a potential of targeting to the membrane or tonoplast. In contrast, no transmembrane domain was predicted by TM finder (Liu and Deber, 1998; Liu and Deber, 1999) (Fig. 3.18) and SUBA proposed other possibilities (Table 3.2).

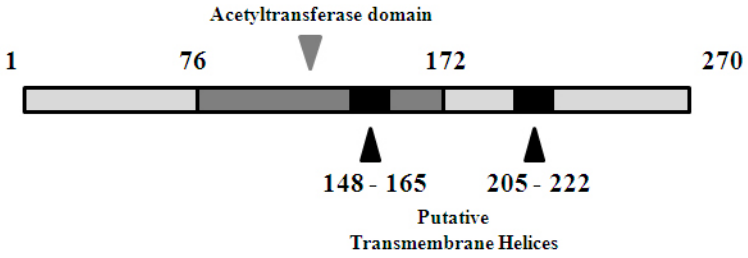


Figure 3.17 Figure 3.5 Schematic model of AtNaa60p.

AtNaa60p is a protein of 270 amino acids with two putative transmembrane helices which locate in the middle (18 amino acids from 147-165) and at the C-teminus (18 amino acids from 205-222). The acetyltransferase domain is between amino acids 147-227. The putative transmembrane helices were predicted by PredictProtein (Rost et al., 2004). SMART was used to analyze the acetyltransferase domain (Schultz et al., 1998; Letunic et al., 2012).

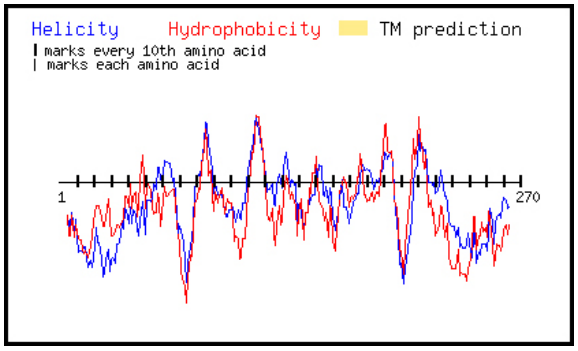


Figure 3.18 AtNaa60p was predicted containing no transmembrane domain by TM finder.

Table 3.2 SUBA prediction of AtNaa60p subcellular localization

Predictors	Subcellular localization
iPSORT	no data
LOCtree	no data
MitoPred	no data
MultiLoc	no data
PeroxP	no data
Predotar	no data
SubLoc	**extracellular**
Mitoprot 2	no data
TargetP	no data
WoLFPSORT	**nucleus**

In order to experimentally investigate where AtNaa60p localize, EYFP was fused to the C-terminus of AtNaa60p and transiently expressed in Arabidopsis protoplasts. *ATNAA60* cDNA was amplified by PCR with primers number 1956 and 2085, the restriction sites *BamHI* and *SalI* were added into primers for cloning into pFF19-EYFP vector (Table S2). The pFF19-EYFP-AtNaa60 and pKF13-TMD23-RFP (RFP-plasma membrane marker) (Table S2) were co-transfected into prepared Arabidopsis protoplasts for co-localization (sections 2.2.6 and 2.2.7).

The localization of AtNaa40-EYFP fused protein and RFP-cytosolic marker was shown in the Fig. 3.18. In agreement with PredictProtein (Rost et al., 2004) prediction, signal of AtNaa60-EYFP fused protein was detected in the plasma membrane (Fig. 3.18A). However, co-localization with RFP-PM marker revealed that only part of AtNaa60p

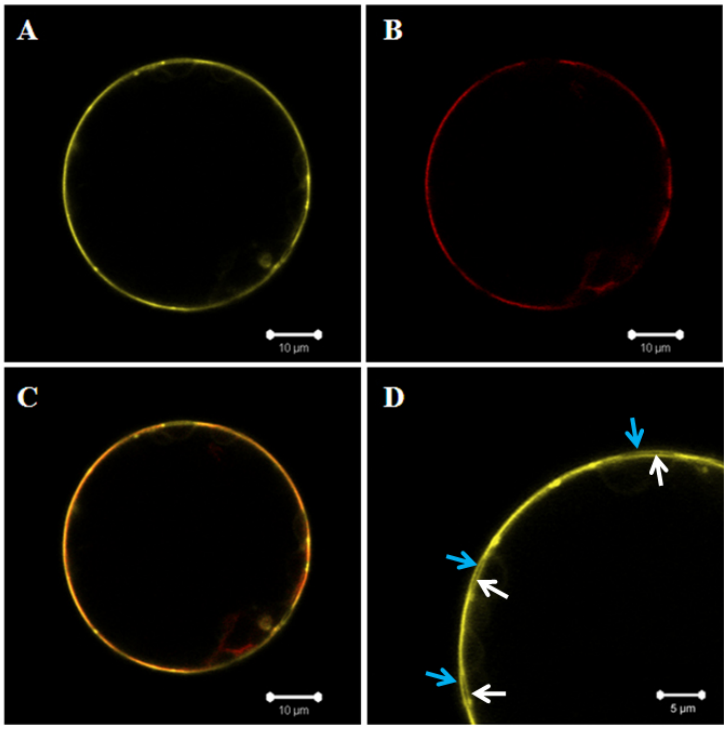


Figure 3.18 Co-localization of AtNaa60-EYFP fused protein and RFP-plasma membrane marker in Arabidopsis protoplasts.

Constructs were transfected into Arabidopsis protoplasts, after 18 hours incubation the transient expressions were analyzed by confocal microscopy. (A) Signal of AtNaa60-EYFP fused protein was detected in PM. (B) TDM23-RFP (Table S2) was used as the PM marker (Brandizzi et al. 2002). (C) Partly overlay of YFP signal and PM marker indicated localization of AtNaa60p in another compartment. (D) AtNaa60-EYFP fused protein localized both to the PM and to the tonoplast. Blue arrows points to PM, white arrows points to tonoplast.

localized to the plasma membrane, another part was present in the membrane-like compartment (Fig. 3.18C). Further analysis of transfected protoplasts showed a trend of AtNaa60p localization. In many protoplasts, interspaces were found between two yellow layers of YFP signals. The ourter layer is plasma membrane and the inner layer (membrane-like compartment) is probably tonoplast (Fig. 3.18D).

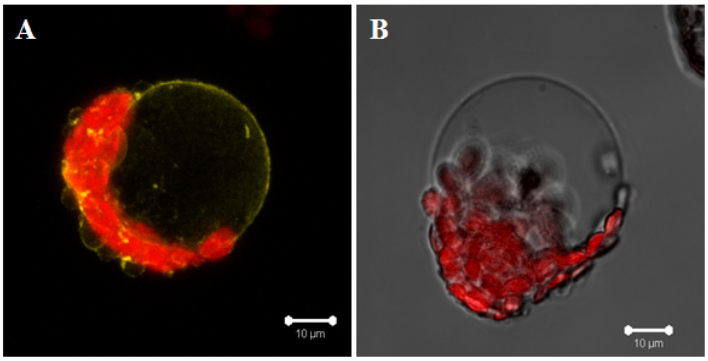


Figure 3.19 Localization of AtNaa60-EYFP fused protein and RFP-PM marker after PM osmotic lysis.

(A) Signal of AtNaa60-RYFP fused protein was still remaining in the tonoplast. (B) No signal of RFP was detected after osmotic lysis of PM. Red colour indicates autofluorescen of chloroplasts.

To eliminate any possible interference in tonoplast localization analysis, protoplasts were treated with osmotic solution to expose tonoplast (section 2.2.8). Osmotic lysis of plasma membrane clearly showed that the membrane-like compartment is indeed the tonoplast (Fig. 3.19A). In the negative control, no signal of RFP-PM marker was detected (Fig. 3.19B).

In higher eukaryote, so far six NAT complexes (NatA-NatF) have been identified. Five of them (NatA-NatE) are located in the cytoplasm, some are also found in the nucleus. Therefore, the localization of AtNaa60 to the plasma membrane and tonoplast is surprising and doubtful. To reinforce this result, YFP was fused with AtNaa60p to C-terminus or N-terminus and transiently expressed in tobacco leaves using gateway constructs. *AtNAA60* cDNA was amplified by PCR using primers number 3125, 3126 and 3127 for cloning into pB7YGW2 and pB7WGY2 vector which express AtNaa60-YFP and YFP-AtNaa60, respectively (Table S2). These constructs were infiltrated into tobacco leaves using *Agrobacterium tumefaciens* strain GV 3101 (section 2.2.9).

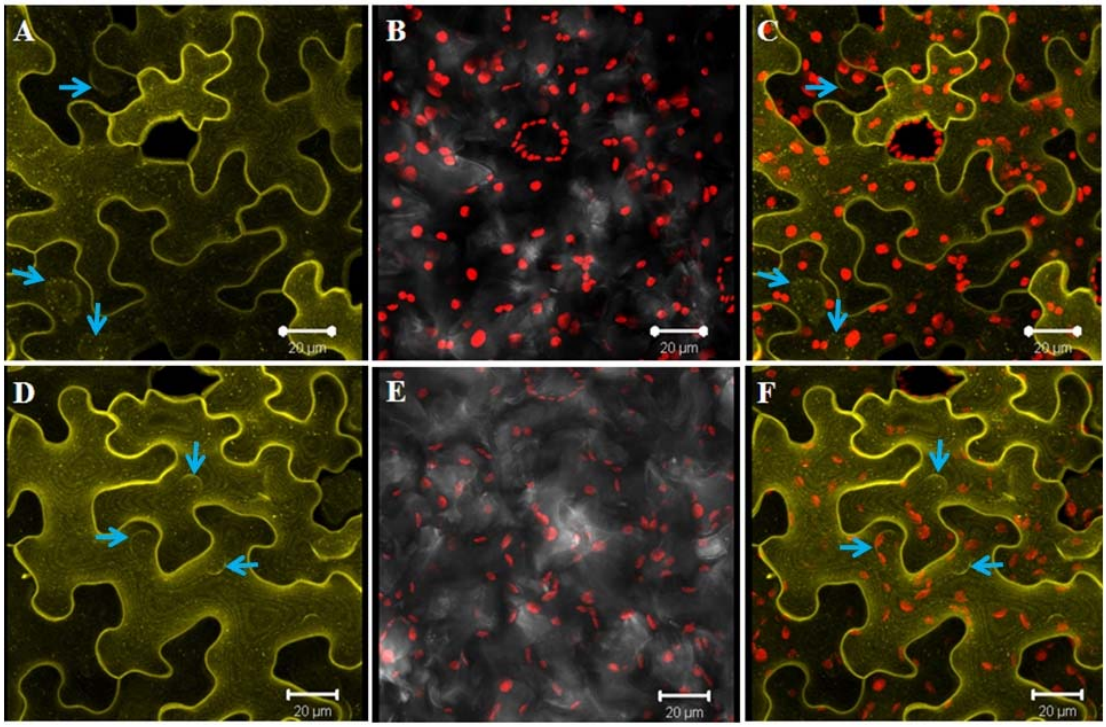


Figure 3.20 Transient expression of YFP-fused AtNaa60p in tobacco leaves.

Agrobacterium tumefaciens carrying pB7YGW2-Naa60 or pB7WGY2-AtNaa60 was infiltrated into tobacco leaves. Two days after infiltration, the expression was analyzed using confocal microscopy. Blue arrow indicates tonoplast-like structure. (A, B, C) Expression of AtNaa60p fused with YFP at the C-terminus. (D, E, F) Expression of AtNaa60p fused with YFP at the N-terminus. (A and D) YFP signal, (B and E) autofluorescence of chloroplasts, (C and F) overlay.

The expression of AtNaa60p fused with YFP at the C-terminus in tobacco leaves showed the same localization when it was expressed in Arabidopsis protoplasts. YFP signal of AtNaa60-YFP fused protein was mostly found in the plasma membrane and partly in the tonoplast-like structure (Fig. 3.20 A, B and C). Besides, N-terminus YFP-fused AtNaa60p did not show any differences in the localization (Fig. 3.20 D, F and F). The unique localization in both cases of N-terminal fusion and C-terminal fusion indicated that the fusion with YFP protein did not affect on the subcellular localization of AtNaa60p. In addition, the co-localization with PM marker in tobacco leaves revealed that AtNaa60p localized not only to the plasma membrane but also to the membrane-like compartment (Fig. 3.21), the same pattern it displayed in Arabisopsis protoplasts.

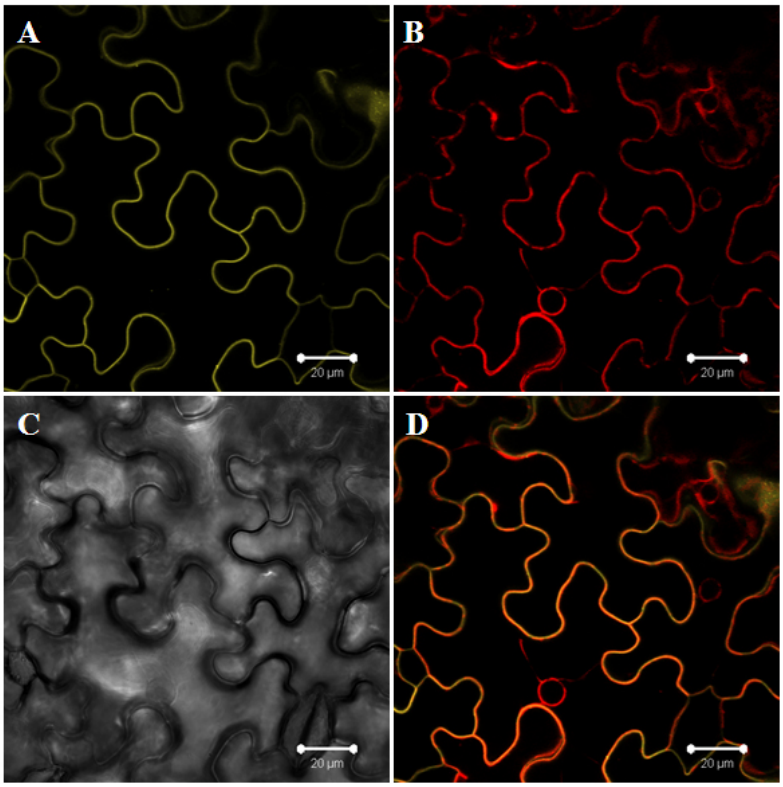


Figure 3.21 Expression of tobacco leaves harbouring p35S:AtNaa60-YFP and pUbi-TDM23-RFP.

Agrobacterium tumefaciens carrying pB7YGW2-AtNaa60 (expresses AtNaa60-YFP) and *Agrobacterium tumefaciens* carrying pUbi-TDM23-RFP were mixed and co-infiltrated into tobacco leaves. Two days after infiltration, the expression was analyzed using confocal microscopy (A) YFP signal. (B) RFP signal. (C) Bright field. (D) Partly overlay of YFP signal and RFP signal indicates the localization of AtNaa60p not only to the plasma membrane but also to the other compartments.

Normally, tonoplast occupies most of the cell space and gets close to plasma membrane and the cell wall (Fig. 3.20 and 3.21), therefore it interferes the localization analysis. To distinguish between the localization in plasma membrane or tonoplast, tobacco leaves were cut into small pieces (1cm x 1cm) and soaked in 0.5 M NaCl solution for one hour. Under high osmotic pressure, the plasma membrane was sunk away and separated from cell wall and tonoplast (Fig. 3.22). As shown in the Fig. 3.22A and 3.22D, two chloroplasts located in the interspace of two separated membranes which are the tonoplast and plasma membrane. The detection of YFP signal in these membranes (Fig. 3.22 C and D) clearly proved the localization of AtNaa60p both to the tonoplast and to the plasma membrane.

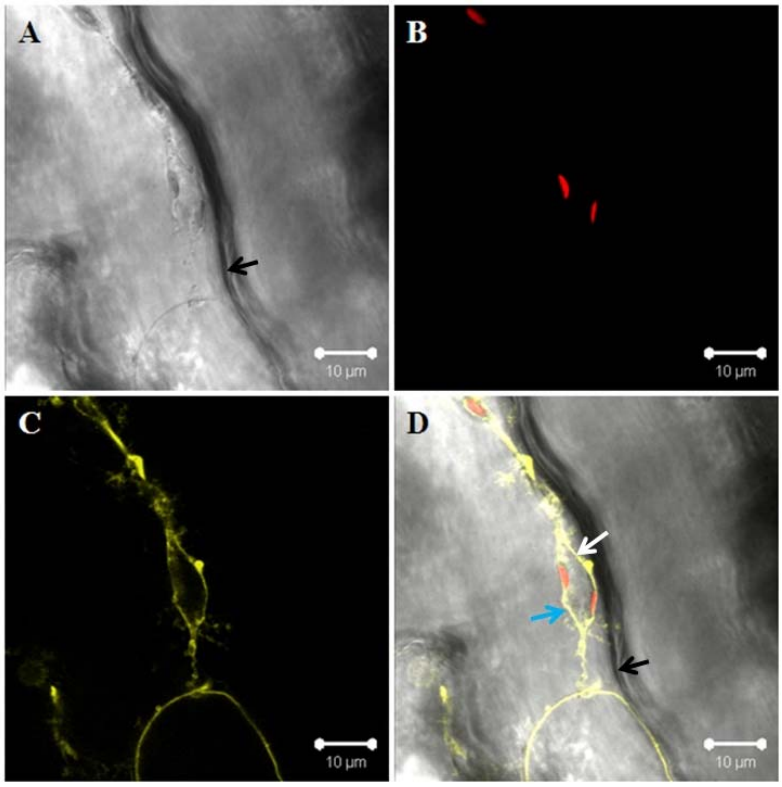


Figure 3.22 Expression of AtNaa60-YFP fused protein in tobacco leaves after osmolysis.

Small pieces (1cm x 1cm) of tobacco leaves expressing AtNaa60-YFP fused protein were treated with 0,5 M NaCl for 1 hour before analyzing using confocal microscopy. Black arrow indicates cell wall, blue arrow indicates tonoplast and white arrow indicates plasma membrane. (A) Bright field. (B) Autofluorescence of chloroplasts. (C) YFP signal. (D) Overlay of A, B and C.

3.2.5 AtNaa60p potentially associates with another membrane-anchored protein

AtNaa60p is a 270 amino acids protein which is predicted to contain two putative transmembrane helices. Transient expression of AtNaa60p fused with YFP in Arabidopsis protoplasts and tobacco leaves reveals the localization both to the tonoplast and plasma membrane (section 3.2.4). However, it was reported that the length of the transmembrane domain plays a role in subcellular localization in plant. Protein with 23 or 22 amino acids transmembrane domain was accumulated in plasma membrane, 20 or 19 amino acids transmembrane domain targeted protein into the Golgi and the shorter transmembrane domain of 17 amino acids led to ER retention (Brandizzi et al., 2002). Two transmembrane domains of AtNaa60p contain only 18 amino acids which are probably not sufficient for targeting to the plasma membrane or tonoplast. An alternative way to target AtNaa60p to

the plasma membrane and the tonoplast would be associated with a membrane-anchored protein.

In order to investigate the possibility of AtNaa60p interacting with a membrane-anchored protein, membrane fraction was isolated from infiltrated tobacco leaves. Protein-protein interactions were disrupted by Na_2CO_3 treatment (section 2.5.7). Fluorescent-tagged proteins from different fractions were detected using GFP-antibody (section 2.5.8). As shown in Fig. 3.23, AtNaa60-YFP was found in the membrane fraction. A small amount of AtNaa60-YFP was also present in the supernatant, which could be the result of disrupting membrane during the membrane isolation procedure. As expected, AtNaa60-YFP was washed away from the membrane by Na_2CO_3 indicating that AtNaa60p possibly associates with an integral membrane protein. AtERDL6-GFP, a vacuolar glucose exporter was used as a positive control. Unfortunately AtERDL6-GFP was not detected in all fractions, probably the amount of protein was not high enough to detect. A cytosolic protein, YFP-SAT5 was used to control any contaminations of soluble proteins in membrane fraction. However, this protein was found both in the supernatant and in the membrane fractions indicating a contamination of the sediment with cytosolic proteins. Besides, the aggregation of over-expressed protein could be a reason why YFP-SAT5 was detectable in the pellet. Taken together, the western-blot result suggested that AtNaa60p possibly associates with an integral membrane protein but the further evidences with suitable controls have to be proven.

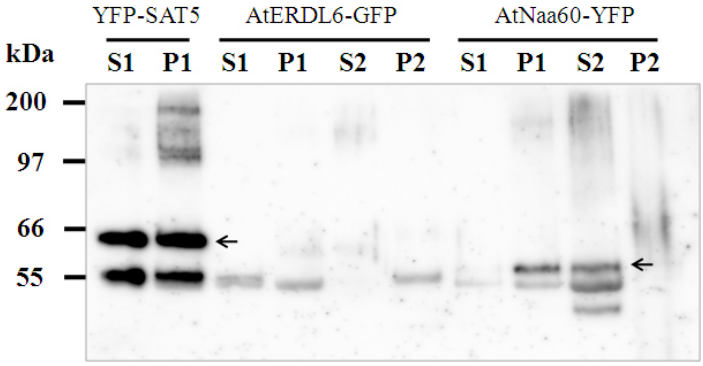


Figure 3.23 AtNaa60-YFP was detected in the Na_2CO_3 extracted-supernatant fraction.

Membrane fraction was washed with 0,1 M Na_2CO_3 to disrupt electrostatic interactions with integral proteins or polar head of lipid. 1 µg of protein was loaded on the gel. GFP-antibody was used to detect the fluorescent tag-proteins in different fractions. (S1) supernatant fraction, (P1) membrane fraction/pellet 1, (S2) Na_2CO_3 extracted-supernatant fraction, (P2) Na_2CO_3 extracted membrane fraction. Black arrow indicates the specific bands. YFP-SAT5, a cytosolic protein was found in S1 and S2. AtERDL6-GFP, a vacuolar glucose exporter was not detected in any fractions. AtNaa60-YFP was washed away from membrane fraction.

3.2.6 Identification and characterization of *atnaa60* mutant

3.2.6.1 Identification of *atnaa60* mutant

To further characterize the significance of NatF in *Arabidopsis thaliana*, a mutant (SALK_016406C) with T-DNA insertion at locus *AT5G16800* was isolated from SALK collection. *AT5G16800* is a gene containing 5 exons, the T-DNA was inserted in the first intron (Figure 3.24). The homozygote plants were identified by PCR genotyping as mentioned in identification of *atnaa40* mutants (section 3.1.5.1). The homozygote mutants

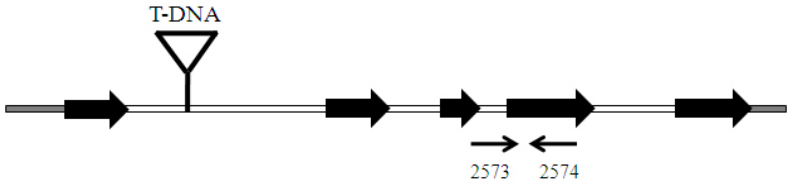


Figure 3.24 Structure of *ATNAA60* with T-DNA insertion.

atnaa60 mutant has T-DNA insertion at the first intron. Big black arrow indicates exon, white bar indicates intron, grey bar indicates un-translated region (UTR) and small back arrow indicates qRT-PCR primers.

with two T-DNA insertions in both alleles had T-DNA specific band, whereas the wild type plant with no T-DNA insertion had gene specific band (Fig. 3.25).The homozygote plants were used for further experiments.

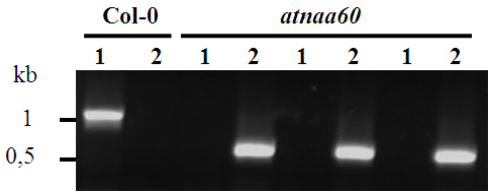


Figure 3.25 PCR genotyping of *atnaa60* T-DNA insertion mutant.

PCR genotyping of *atnaa60* mutant was run with two pairs of primers. (1) Gene specific primers: 2425 and 2426. (2) T-DNA insertion specific primer: 432 and 2426.

3.2.6.2 The expression of *ATNAA60* is severely disrupted in *atnaa60* mutant

To evaluate the effect of T-DNA insertion on the function of *ATNAA40* gene, qRT-PCR with primers number 2573 and 2574 (Fig. 3.24) was performed (section 2.4.8). Similar to *atnaa40-2* mutant, the T-DNA insertion at the first intron severely disrupted the expression of *ATNAA60.* In *atnaa60* mutant, the mRNA steady state level of *ATNAA60* was only about 1.6 % remanining (Fig. 3.26).

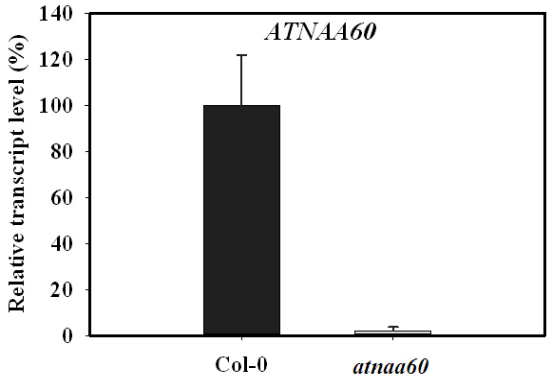


Figure 3.26 Expression of ATNAA60 in the T-DNA insertion mutant.

qRT-PCR was run with primers 2573 and 2574. T-DNA insertion at the first intron severely disrupted the expression of *ATNAA60* in the mutant. Error bar indicates standard error (n=3).

3.2.6.3 Phenotypes of *atnaa60* mutant

Human Naa60p was reported to contribute about 10 percent of protein N-terminal acetylation. Knockdown or overexpression of hNAA60 in HeLa cells was found to increase or decrease the acetylation of protein respectively. In Drosophila, depletion of d*NAA60* resulted in chromosomal defects during anaphase (Van Damme et al., 2011c). In *Arabidopsis thaliana*, at the individual level, the functional disruption of *ATNAA60* did not affect the phenotypes of the mutant. Under normal growth conditions (section 2.2.2), *atnaa60* mutant displayed the normal growth, normal development and ability of producing seeds (Fig. 3.27).

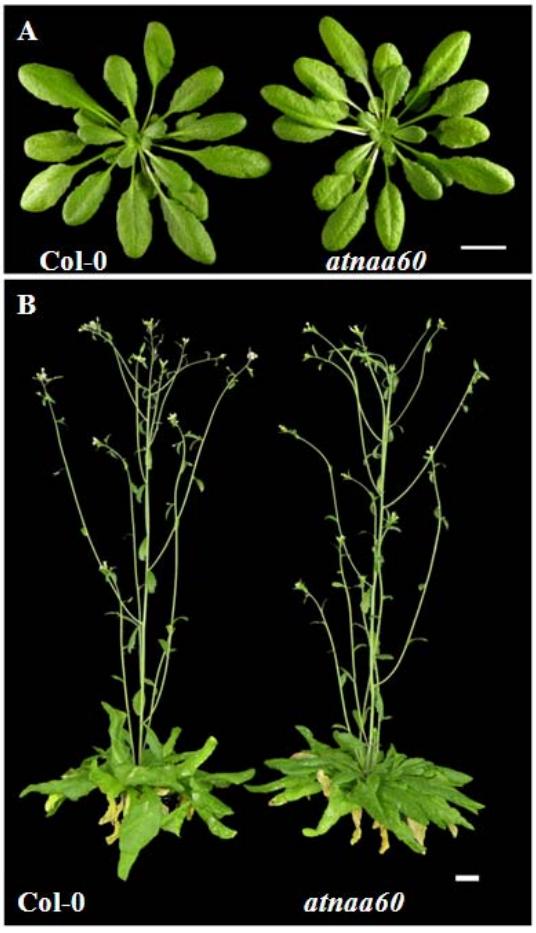


Figure 3.27 Phenotypes of *atnaa60* mutant under normal conditions.

Seeds were planted in the soil under short day conditions, after 10 weeks, plants were transferred to the long day chamber for producing seeds. (A) Seven week-old plants under short day conditions. (D) Eleven week-old plants were flowering and producing seeds in the long day chamber. White bar indicates scale of 2cm.

2.3.6.4 Germination of *atnaa60* seeds in different media

Unlike the other NATs, AtNaa60p localizes to the plasma membrane and to the tonoplast (section 3.2.4) suggesting another function beside Nt-acetylation. Cell membrane separates

the interior of all cells from the outside environment. It is involved in variety of cellular processes such as cell adhesion, ion conductivity and cell signaling. Plant vacuole occupies 30% to 80% of the total volume of the cell and one of its major roles is turgor maintenance. Vacuolar membrane contains numerous proteins that are responsible in the transport of small solutes and salts and proteins that are involved in the fusion of membrane and remodeling the allowed delivery of macromolecules by vesicle traffic (Marty, 1999).

In order to investigate the role of AtNaa60p in the osmotic stress response, seeds of *atnaa60* mutant were germinated on the MS media supplemented with either NaCl or mannitol. On the MS media, seeds from wild type or *atnaa60* mutant germinated normally (about 90% of seeds germinated). However, under high salt stress (100mM NaCl), only 50% of *atnaa60* seeds were able to germinate, whereas more than 70% of wild type seeds germinated successfully. At the higher concentration of NaCl (120mM), the *atnaa60* seeds struggled with germination (Fig. 3.28A). Surprisingly, there was no significant difference in the germination of seeds between the wild type and *atnaa60* mutant when they were germinated on the MS media or MS media supplemented with 165 mM mannitol (Fig. 3.28B). These results indicated that the germination of *atnaa60* seeds was specifically affected under salt stress but not osmotic stress.

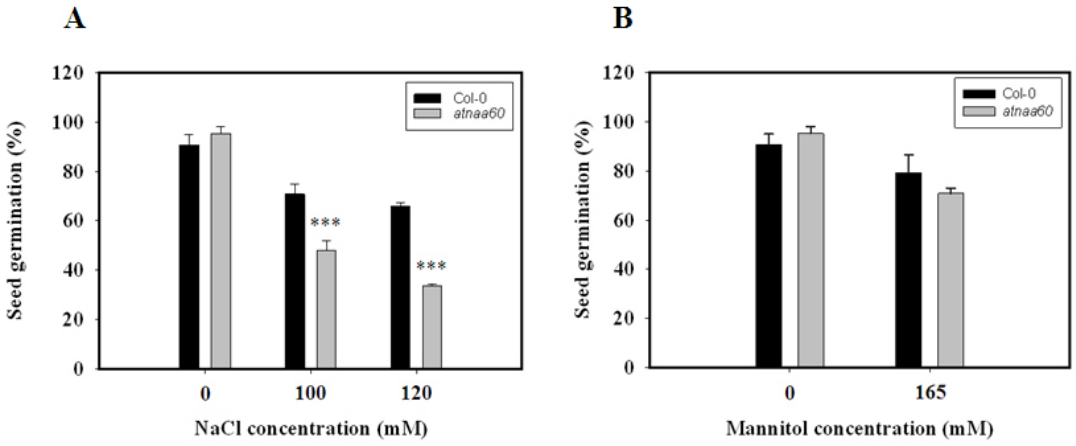


Figure 3.28 Germination of atnaa60 seeds on the high osmotic pressure media

Seeds from *atnaa60* plants and the wild type were sterilized and germinated on the MS media or MS media supplemented with either NaCl or mannitol. Two weeks after plating, the seed germination was counted. Error bar indicates standard error (n=3). Three star indicates the significant difference ($p < 0{,}001$). (A) Percents of seed germination on the MS media supplemented with 100 mM or 120 mM NaCl. (B) Percents of seed germination on the MS media supplemented with 165 mM mannitol.

2.3.6.5 Phenotypes of *atnaa60* seedlings under stress conditions

The functional disruption of *ATNAA60* resulted in the sensitivity of *atnaa60* mutant to salt stress during germination of seeds. In order to further investigate the impact of salt tress and osmotic stress to *atnaa60* mutant during development, sterilized seeds were germinated on the MS media. Five days after planting on the places, germinated seeds were transferred to the MS media supplemented with either NaCl or mannitol. Three week-old seedlings on different media were shown in the Fig. 3.29. Seedlings under salt stress condition grew slowly and had darker leaves. Under osmotic stress, they grew even more slowly and had curve leaves as the result of lack of water. However, no difference was observed between the wild type and the *atnaa60* mutant. Furthermore, there was no significant increase in the expression of *ATNAA60* in seedlings grown on the salt containing media, whereas the expression of *DREB2A*, a marker for salt stress was increased 7-fold (Fig. 3.30). Taken together, AtNaa60p in some ways has a role in salt stress response at the germination state since *atnaa60* mutant was sensitive to salt stress during seed germination. However, this function seems to change during plant development.

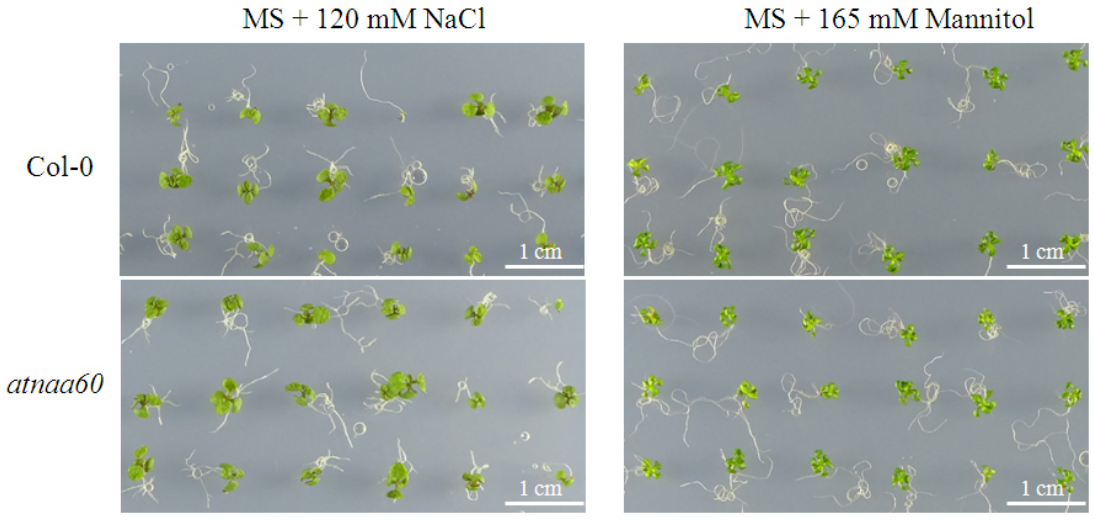


Figure 3.29 Three week-old seedlings of wild type and *atnaa60* mutant on MS media supplemented with 120 mM NaCl or 165 mM mannitol.

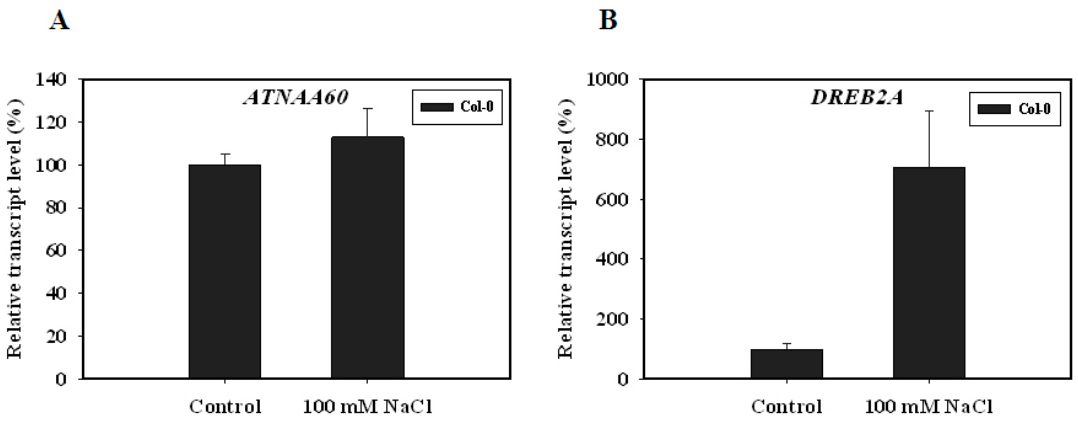


Figure 3.30 Expression of *ATNAA60* and *DREB2A* under salt stress condition.

cDNA synthesized from total RNA of 3week-old seedlings growing on the MS medium supplemented with 100 mM NaCl was used for qRT-PCR analysis. Error bar indicates standard error (n = 3). (A) No significant increase in the expression of *ATNAA60*. (B) As a marker for salt stress condition (Sakuma et al., 2006), the expression of *DREB2A* increased 7-fold.

2.3.6.6 Status of protein N-terminal acetylation in the *atnaa60* mutant

NatA, NatB and NatC are three major NATs that N-terminally acetylate most of proteins in the eukaryotic cells. hNatF, a novel type of NAT, contributes about 10% of protein acetylation (Van Damme et al., 2011c). In order to assess the contribution of AtNaa60p to overall protein acetylation in plant, unacetylated protein level in *atnaa60* mutant was quantified using NBD-Cl fluorescent assay (section 2.5.6). As shown in the Fig. 3.31, the fluorescent intensity in *atnaa60* mutant sample was 2-fold higher than the wild type indicating a higher level of unacetylated proteins. The high level of unacetylated protein in the *atnaa60* mutant revealed a significant quantity of proteins was acetylated by AtNaa60p. This result was expected because hNaa60p acetylates about 10% of proteins and AtNaa60p possibly has even more substrates than hAtNaa60p. However, since the *atnaa60* mutant did not show any phenotypes, the role of protein acetylation by AtNaa60p is remaining unclear.

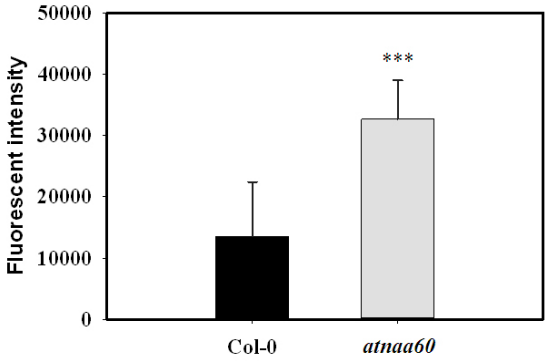


Figure 3.31 **Status of protein N^{α}-acetylation in the *atnaa60* mutant.**

7,6 µg of purified protein extract was reacted with 0.5 mM NBD–Cl in 50 mM sodium citrate buffer overnight at RT. Overnight reaction solution was exited at 470 nm and fluorescence intensity was measured at 520 nm. Error bar indicates standard error (n=4).

3.3 Identification of plastidic N^{α}-acetyltransferases (pNATs)

3.3.1 Searching for the putative pNATs

Recently, it was reported that abundant proteins were N-terminally acetylated in the chloroplast (Kleffmann et al., 2007; Zybailov et al., 2008; Bischof, 2010; Bienvenut et al., 2012). The N^{α}-acetyltransferase and other member of GNAT family proteins are characterized by for conserved regions (motif A-D) spanning more than 100 residues (Neuwald and Landsmann, 1997). In a search of novel plastidic NATs, all proteins that contain GNAT domain were considered. In total of 74 loci listed on TAIR, those genes coding for putative histone acetyltransferases or other acetyltransferases were eliminated. Eight putative NATs that are predicted to localize in the chloroplast were listed in Table 3.3.

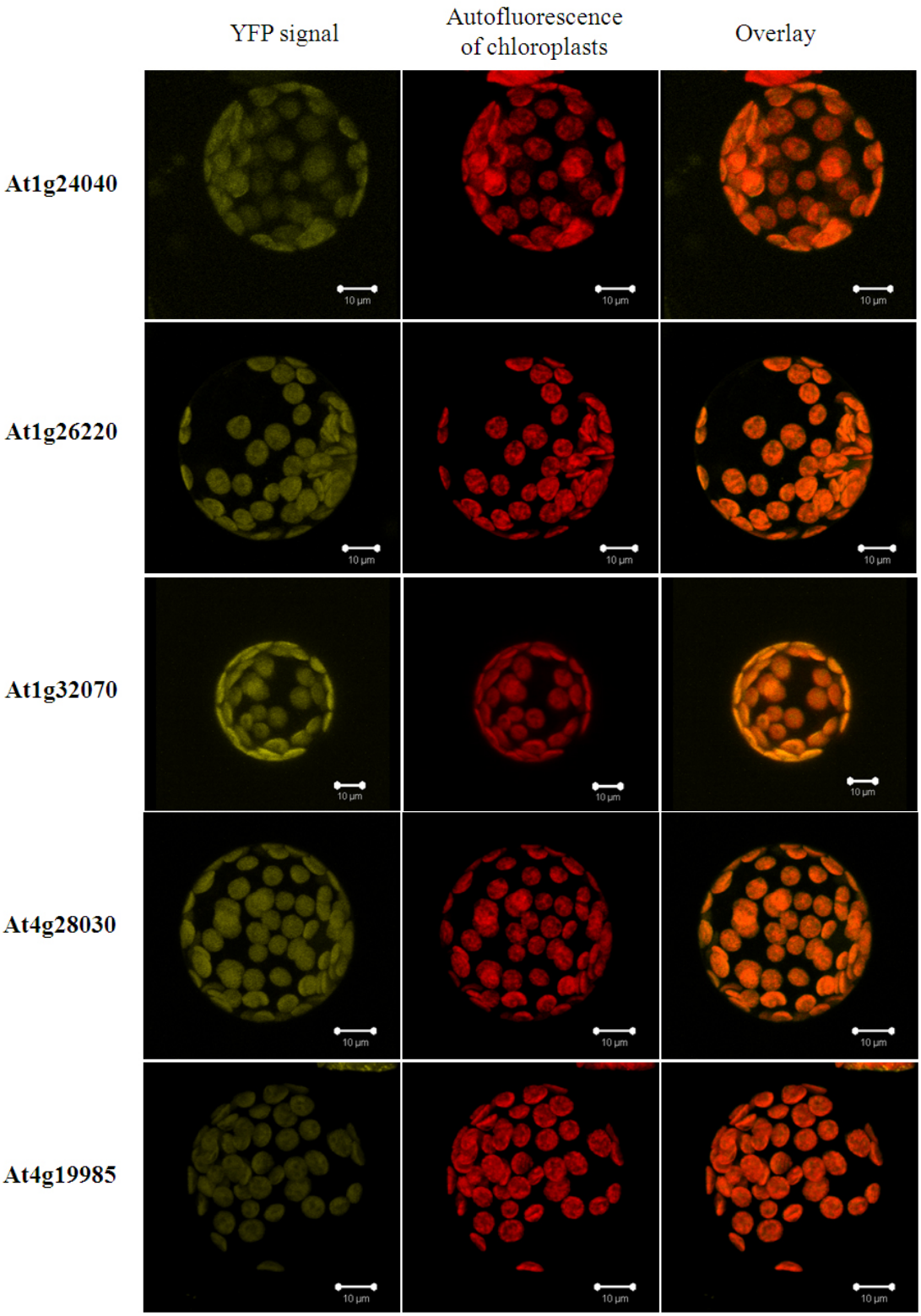
Table 3.3 Putative plastidic N^{α}-acetyltransferases

No.	Locus	Mainly predicted localization by SUBA
1	At1g24040	Plastid
2	At2g06025	Plastid, Mitochondria
3	At4g28030	Plastid
4	At2g39000	Plastid
5	At4g19985	Plastid, Mitochondria
6	At1g26220	Plastid
7	At1g32070	Plastid
8	At1g72030	Plastid, Mitochondria

3.3.2 Subcellular localization of putative pNATs

In order to address in which subcellular compartment the putative NATs localize, these proteins were fused with EYPF at the C-terminus and transiently expressed in Arabidopsis protoplasts. cDNA of putative NATs were amplified by PCR using specific primers containing restriction site (listed in section 2.1.6) for cloning into pFF19-EYFP (Table S2).These constructs were transfected into prepared Arabidopsis protoplasts for localization analysis (section 2.2.6 and 2.2.7). Transient expression of EYFP fused proteins in Arabidopsis protoplast confirmed the localization of seven candidates to the

chloroplasts. Another candidate, At1g72030 was likely targeted to the mitochondria (Fig. 3.32).



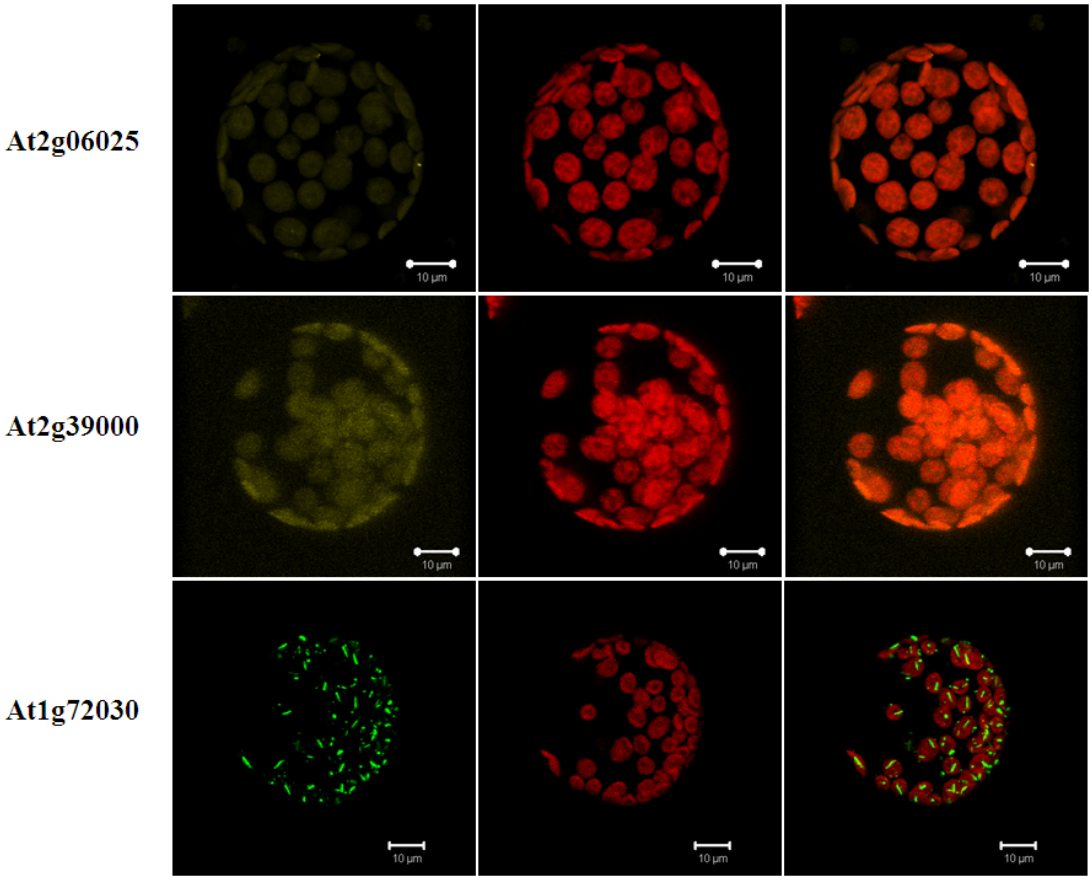


Figure 3.32 Subcellular localization of putative pNATs in Arabidopsis protoplasts.

cDNA of putative plastidic NATs were amplified and cloned into pFF19-EYFP. Constructs were transfected into Arabidopsis protoplasts. After 18 hours incubation, the transient expression of EYFP fused proteins were analyzed using confocal microscopy. In case of At1g72030, yellow fluorescence was indicated in green.

3.3.3 *In silico* analysis of putative pNATs

In the evaluation of seven putative pNATs whose subcellular localizations were confirmed by YFP fusion, their amino acid sequences were used for alignments and creating a phylogenetic tree. As shown in Fig. 3.33, six candidates except At4g28030 contain Ac-CoA binding motif (Q/RxxGxG/A) which is essential for activity of acetyltransferases. The existence of conserved Ac-CoA binding motif strongly indicates these candidates as the true NATs. Although it was reported that the first amino acid of Ac-CoA binding motif can be either G or R (Polevoda et al., 1999), R residue seems to be the first priority. In all known NATs from yeast, human and Arabidopsis, only AtNaa40p accepts Q as the first

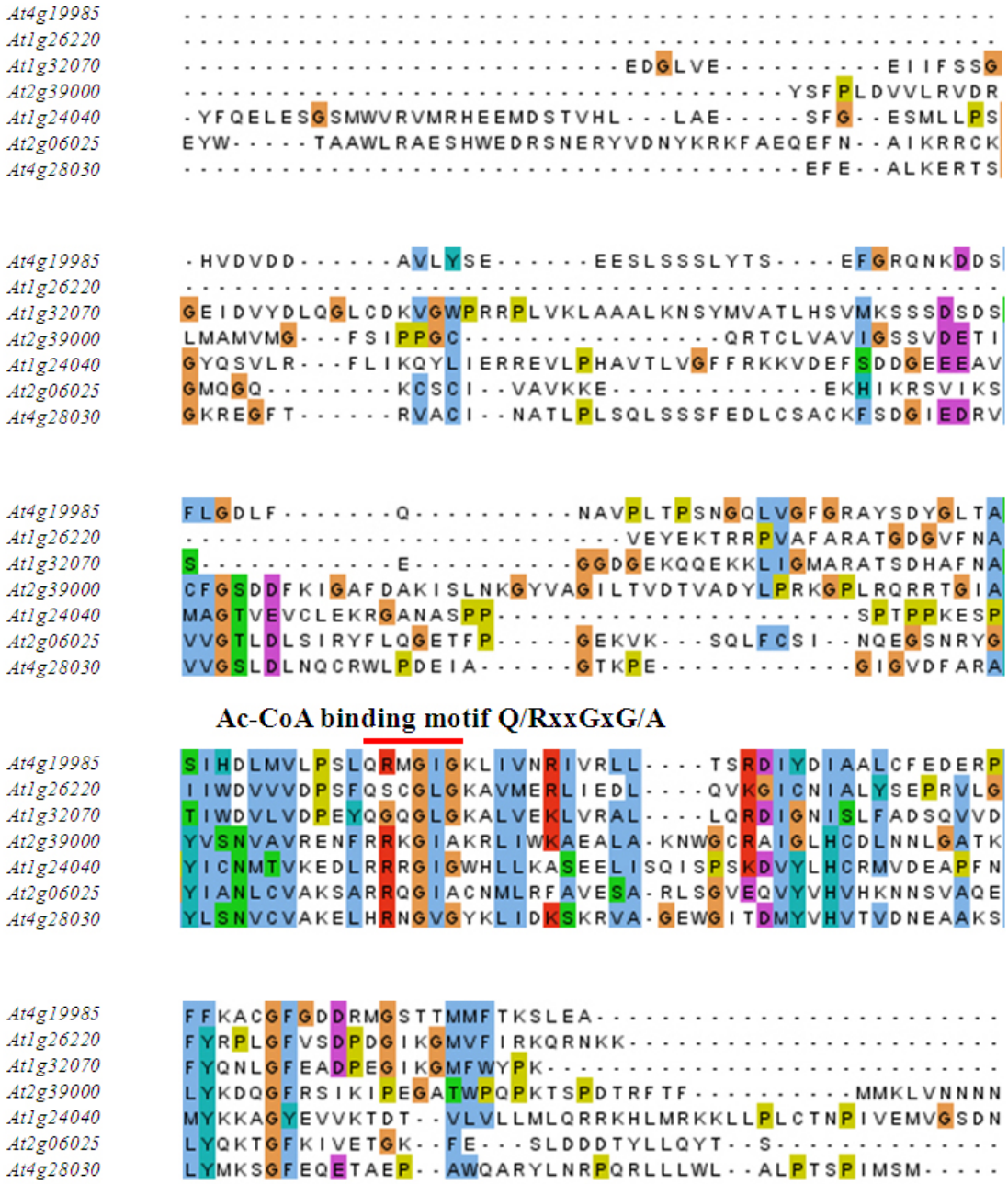


Figure 3.33 Amino acid sequence alignments of putative plastidic NATs.

Amino acid sequences (starting from 101 residue) of putative plastidic NATs were aligned using Clustal Omega (Goujon et al., 20010; Sievers et al., 2011). The conserved Ac-CoA binding motif Q/RxxGxG/A, where x represents any amino acid is marked with the red line above. The red background indicates basic residues, purple indicates acid residues, orange indicates proline, blue indicates hydrophobic residues, green indicates polar residues, and turquoise indicates histidine and tyrosine residues. Amino acids were highlighted using Jalview (Waterhouse et al., 2009).

amino acid in Q/RxxGxG/A sequence. Other NATs employ R residue in their sequences. Like most of NATs, three putative pNATs (At2g39000, At1g24040 and At2g06025) possess R residue in Ac-CoA binding motif. In contrast, three other putative pNATs (At4g19985, At1g16225, At1g32070) use Q residue instead. Noticeably, further analysis revealed that At2g39000, At1g24040 and At2g06025 belong to three different subgroups (Fig. 3.34). Since At2g39000, At1g24040 and At2g06025 present for three different subgroups and their Ac-CoA binding motifs contain preferential R residue, they are selected for further analyses to look for the N^{α}-acetyltransferases that are responsible for Nt-acetylation in the chloroplast.

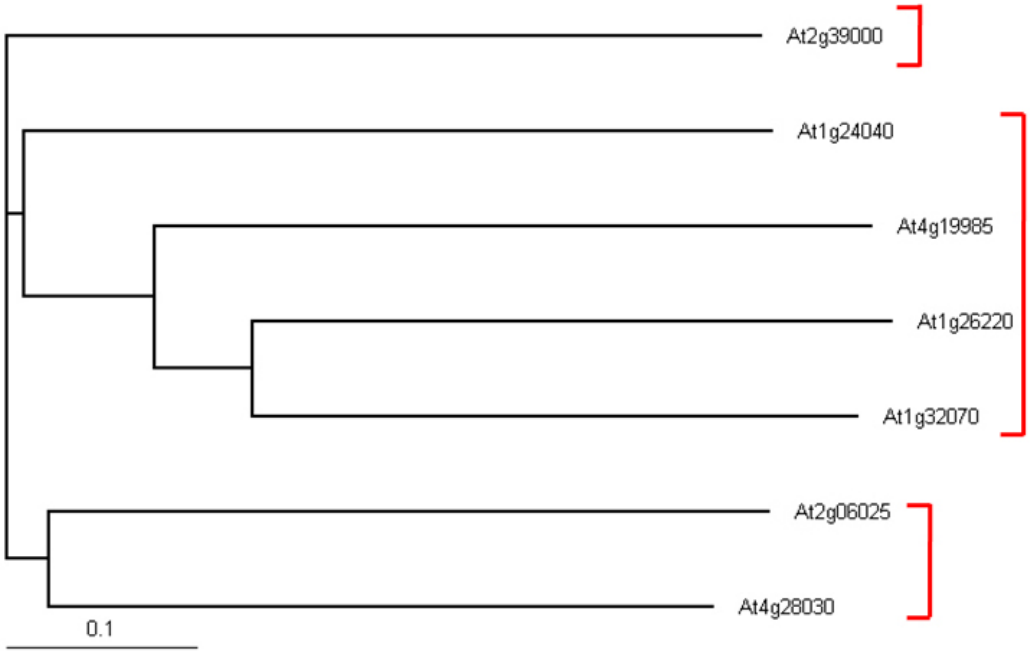


Figure 3.34 Phylogenetic tree of seven putative plastidic NATs.

Clustal Omega was used to create the phylogenetic tree from seven plastidict NATs. Phylogenetic tree was viewed by TreeView (Page, 1996). Red square bracket indicates different subgroups.

3.3.4 Overexpression of putative pNATs in *E. coli*

In the purpose of preparation for substrate specificity analysis, recombinant putaitive NATs proteins were expressed in *E. coli*. cDNAs of At2g06025, At2g39000 and At1g24040 were amplified by PCR with specific primers containing restriction sites (listed in section 2.1.6) for cloning into pETM41 expression vector (Table S2). Since the transit

peptide might have effects on the function, it was cleaved off before cloning. At2g06025, At2g39000 and At1g24040 were predicted to contain 111, 61 and 55 amino acids transit peptide respectively by TargetP (Nielsen et al., 1997; Emanuelsson et al., 2000). The expressions of pETM41- At2g06025, pETM41- At2g39000 and pETM41- At1g24040 in *E. coli* Rosetta strain were induced by 1 mM IPTG overnight at 21^{o}C. The increasing quantity of these proteins displayed time dependent manner indicating the successful expression in *E .coli* (Fig. 3.35).

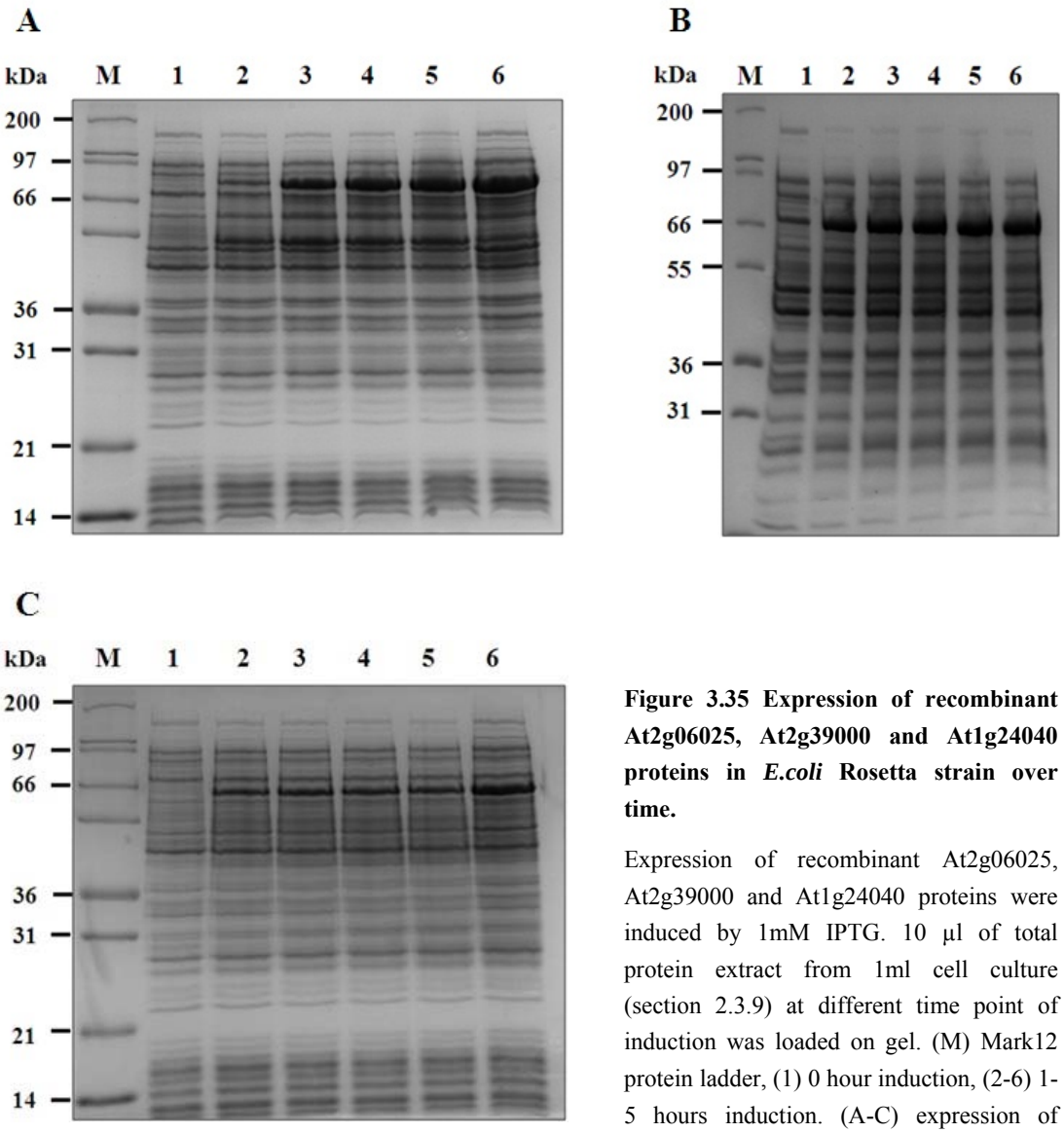


Figure 3.35 Expression of recombinant At2g06025, At2g39000 and At1g24040 proteins in *E.coli* Rosetta strain over time.

Expression of recombinant At2g06025, At2g39000 and At1g24040 proteins were induced by 1mM IPTG. 10 µl of total protein extract from 1ml cell culture (section 2.3.9) at different time point of induction was loaded on gel. (M) Mark12 protein ladder, (1) 0 hour induction, (2-6) 1-5 hours induction. (A-C) expression of At1g24040p, At2g39000p and AT2g06025p.

3.3.5 LC-MS/MS analysis of Nt-acetylation by pNATs

In order to assess whether three putative pNATs (At2g06025, At2g39000 and At1g24040) were the true NATs, a novel approach was conducted to identify substrate specificity of N^{α}-acetyltransferases *in vivo*. The mature putative pNATs were recombinantly expressed in *E. coli* (section 3.3.4) whose proteins are rarely acetylated. In theory, the recombinant pNATs were able to acetylate free N-termini of *E. coli* proteins. The acetylated N-termini were subsequently identified by LC-MS/MS (section 2.5.9). The *in vivo* substrate specificity profile was analyzed using WebLogo (Crooks et al., 2004) and IceLogo (Colaert et al., 2009). In addition, *E.coli* which expressed recombinant AtNaa15p, an auxiliary subunit of AtNatA, was used as the control. All of substrate specificity analysis by LC-MS/MS was done in collaboration with Carmela Giglione's lab, Institute of Plant Science, French National Centre for Scientific Research, Gif/Yvette c, France.

3.3.5.1 Nt-acetylation by At2g39000p (pNaa10p)

a. Identification of recombinant At2g39000p in the sample

In order to prove the presence of At2g39000p in the sample and ensure the additional Nt-acetylations were catalyzed by its activity but not by any bacterial NATs, At2g39000p was identified by mass spectrometry (MS). As shown in Fig. 3.37 and Table 3.4, different fragments of At2g39000p were recognized with 62% of amino acid sequence coverage in total. Based on an exponentially modified protein abundance index (emPAI) of 23,62 out of 2116,5 overall ∑emPAI, the amount of At2g39000p in the lysate was esmimated about 1.1%. Interestingly, four lysine residues (217K, 243K, 245K and 265K) were found acetylated. It could be the result of autoacetylation of At2g39000p or by acetylation of *E. coli* lysine acetyltransferases.

```
  1 MRSTPLGTTA VSPASIKHNS YGYGRFVSSS GVSNFSIHRR RRHSSFSISQ
 51 APSQINSGAC NASQIVDLFP AVSPEIVVRE ARLEDCWEVA ETHCSSFFPG
101 YSFPLDVVLR VDRLMAMVMG FSIPPGCQRT CLVAVIGSSV DETICFGSDD
151 FKIGAFDAKI SLNKGYVAGI LTVDTVADYL PRKGPLRQRR TGIAYVSNVA
201 VRENFRRKGI AKRLIWKAEA LAKNWGCRAI GLHCDLNNLG ATKLYKDQGF
251 RSIKIPEGAT WPQPKTSPDT RFTFMMKLVN NNNTQALEQF R
```

Figure 3.37 identification of At2g39000p by MS.

At2g39000p in sample was identified by MS with 62% amino acid sequence coverage. Matched peptides were highlighted in red. Acetylated lysine residues were shown in bold.

Table 3.4 Identified sequences of At2g39000p by MS

Start	End	Mascot Score	Expect	Peptide
61	79	70	-	C.NASQIVDLFPAVSPEIVVR.E + Deamidated (NQ, N-term)
83	110	63	5.50E-07	R.LEDCWEVAETHCSSFFPGYSFPLDVVLR.V
114	129	46	3.50E-05	R.LMAMVMGFSIPPGCQR.T
160	182	83	1.10E-08	K.ISLNKGYVAGILTVDTVADYLPR.K
190	202	73	9.60E-08	R.RTGIAYVSNVAVR.E
207	213	29	0.0012	R.RKGIAKR.L
214	228	78	3.20E-08	R.LIWKAEALAKNWGCR.A
214	228	49	3.40E-05	R.LIWKAEALAKNWGCR.A + Acetyl (K)
229	251	59	4.20E-06	R.AIGLHCDLNNLGATKLYKDQGFR.S
229	251	33	0.0016	R.AIGLHCDLNNLGATKLYKDQGFR.S + Acetyl (K)
244	251	40	0.00014	K.LYKDQGFR.S
252	271	55	9.70E-06	R.SIKIPEGATWPQPKTSPDTR.F + 2 Acetyl (K)
252	271	56	6.70E-06	R.SIKIPEGATWPQPKTSPDTR.F + Acetyl (K)
255	271	67	3.80E-07	K.IPEGATWPQPKTSPDTR.F + Acetyl (K)
272	291	113	1.10E-11	R.FTFMMKLVNNNNTQALEQFR.-

b. Substrate specificity of At2g39000p (pNaa10p)

Using bacterial LC-MS/MS approach, in total of more than 300 identified N-termini, 182 N-termini were more higher acetylated in the At2g39000p expressing *E. coli*. The acetylation yield was distributed from 6% to 100%, 56 N-termini (31%) were fully acetylated (Table S1B). Of these 182 acetylated N-termini, 75 were starting with methionine, 107 started with other residues. Interestingly At2g39000p substrates covered all substrates of known NATs with preference towards NatA type and NatB type. Besides, N-termini Met-His- and Met-Tyr- so far were not reported to be acetylated by any NATs (Fig. 3.38).

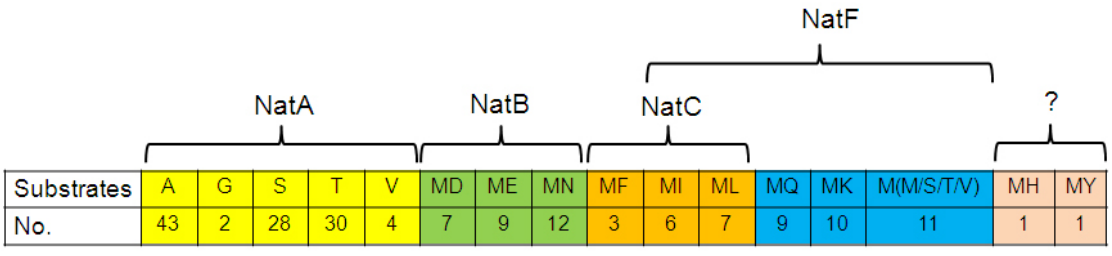


Figure 3.38 At2g39000p substrates identified by LC-MS/MS.

182 substrates of At2g39000p were identified by LC-MS/MS from At2g39000p expressing *E. coli*. Substrate type of NatA and NatB are predominant. MH- and MY- are unique substrates of At2g39000p.

Remarkably, the comparison of At2g39000p substrate specificity versus chloroplast transit peptide (cTP) cleavage sites showed strong correlation between two data sets except the higher frequency of Met that is much more abundant in *E. coli* than in the cTP cleavage sites and the lower frequency of Val that is infrequent in protein N-termini of *E. coli* proteins. The correlation between At2g39000p substrate specificity and cTP cleavage sites strongly reinforces At2g39000p as a plastidic N^{α}-acetyltransferase. At2g39000p was the first identified plastidic N^{α}-acetyltransferase, thus according to the revised NAT-nomeclature system, we named this protein pNaa10p and its activity pNatA.

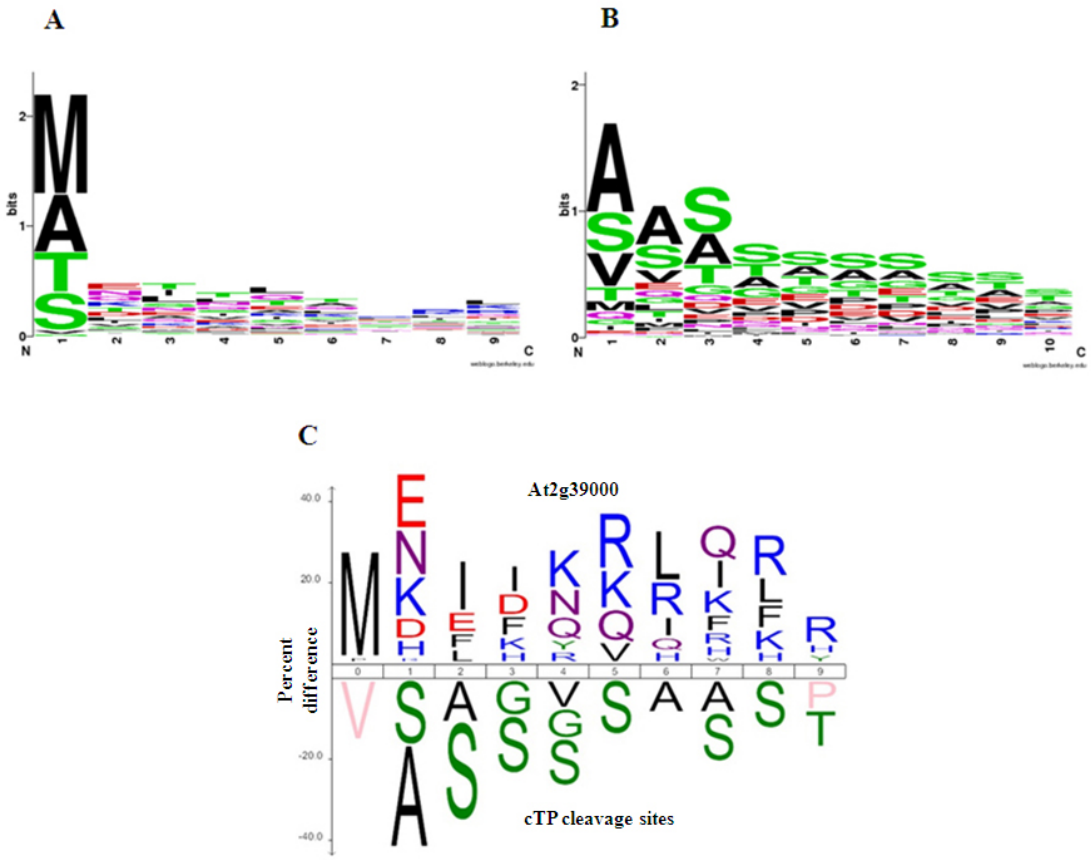


Figure 3.39 Comparison of At2g39000p substrate specificity vs. cTP cleavage sites. (A) WebLogo of At2g39000p substrate specificity (n=182). (B) WebLogo of the cTP cleavage sites related to the experimentally characterized N^{α}-acetylated chloroplastic protein (Bienvenut et al., 2012). Bits indicate the sequence conservation at amino acid position, $\log_2 20 = 4.3$ bits (C) IceLogo for the comparision of At2g39000p substrate specificity *vs*. cTP cleavage site.

c. pNaa10p acetylates lysine residues of itself

As mentioned above, four lysine residues (K217, K243, K254 and K256) were found acetylated when pNaa10p was expressed in *E. coli*. In order to assess the possibility of pNaa10p acetylating lysine residues itself, purified MBP-pNaa10p was used for *in vitro* autoacetylation assay (section 2.5.5). Indeed, the increasing of lysine acetylation signal displayed the time dependent manner indicating the ability of pNaa10p to acetylate its lysine residues (Fig 3.40).

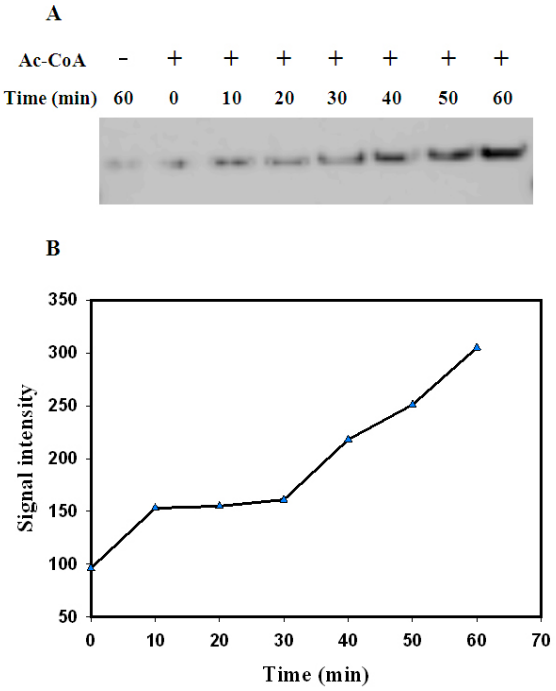


Figure 3.40 Auto-N$^{\varepsilon}$-acetylation of pNaa10p.

Purified MBP-pNaa10p (Fig. S3) was incubated with Ac-CoA for different period of times. After reaction,100 ng of MBP-pNaa10p at different time points was used for western-blot, acetylation of lysine was detected using acetylated-lysine antibody. (A) Western-blot detection of autoacetylation of MBP-pNaa10p. (B) Graph of signal intensity during the time.

3.3.5.2 Nt-acetylation by At1g24040 (pNaa20p)

a. Identification of recombinant At1g24040p in the sample

Like pNaa20p, the presence of At1g24040p in the sample was confirmed by LC-MS/MS. Ten different fragments of At1g24040 were recognized with total amino acid sequence coverage of 57%. Based on the emPAI of 18,26 and the total ∑emPAI for the experiment of 2943, the amount of At1g24040p in the lysate was estimate about 0,6%. However, none of lysine residues was found to be acetylated (Fig. 3.41 and Table 3.5).

```
  1 MAALSISLAF SVDSLKPTQS TKFGFSSSSH RYPLLYSCKS HRSNLRFAFP
 51 PSSVSTATET GEENSKSTGN YAFLEESFRT GRFLSNDELE KLKTLEGFAY
101 FQELESGSMW VRVMRHEEMD STVHLLAESF GESMLLPSGY QSVLRFLIKQ
151 YLIERREVLP HAVTLVGFFR KKVDEFSDDG EEEAVMAGTV EVCLEKRGAN
201 ASPPSPTPPK ESPYICNMTV KEDLRRRGIG WHLLKASEEL ISQISPSKDV
251 YLHCRMVDEA PFNMYKKAGY EVVKTDTVLV LLMLQRRKHL MRKKLLPLCT
301 NPIVEMVGSD NELTSSANV
```

Figure 3.41 Identification of At1g24040 by MS.

At1g24040p in sample was identified by MS with 57% amino acid sequence coverage. Matched peptides were highlighted in red.

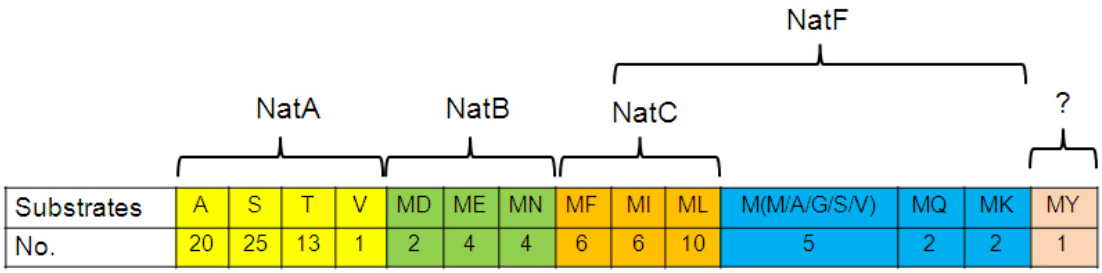
Table 3.5 Identified sequences of At1g24040p by MS.

Start	End	Mascot Score	Expect	Peptide
67	79	63	4.90E-07	K.STGNYAFLEESFR.T
116	145	56	4.20E-06	R.HEEMDSTVHLLAESFGESMLLPSGYQSVLR.F
146	155	54	4.10E-06	R.FLIKQYLIER.R
156	170	55	5.80E-06	R.REVLPHAVTLVGFFR.K
171	197	103	6.50E-11	R.KKVDEFSDDGEEEAVMAGTVEVCLEKR.G
198	226	49	4.20E-05	R.GANASPPSPTPPKESPYICNMTVKEDLRR.R
211	225	52	6.70E-06	K.ESPYICNMTVKEDLR.R
228	255	55	9.20E-06	R.GIGWHLLKASEELISQISPSKDVYLHCR.M
287	292	26	0.0049	R.RKHLMR.K + Oxidation (M)
293	319	77	7.80E-08	R.KKLLPLCTNPIVEMVGSDNELTSSANV.-

b. Substrate specificity of At1g24040p (pNaa20p)

To identify substrate specificity of At1g24040p, more than 230 N-termini were quantified. Of these, 101 N-termini were more acetylated in the At1g24040p expressing *E.coli*. The acetylated yield was distributed between 9% to 100%, only 20 N-termini (20%) were fully acetylated (Table S1C). At1g24040p was able to acetylate both initial methionine (42

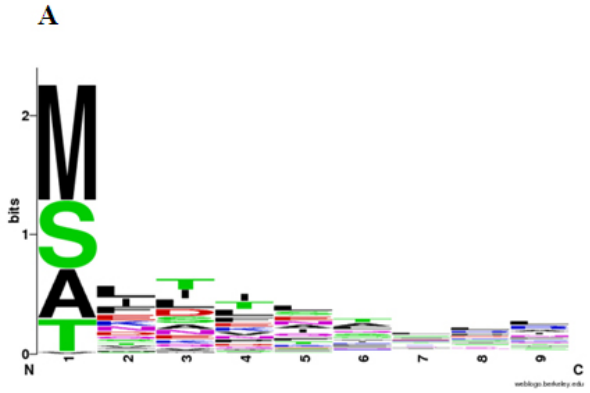
substrates) and the penultimate amino acid residue after removal of methionine (59 substrates). Very similar to pNaa10p, substrates of At1g24040p covered all substrates of all known NATs but preferred towards NatA type and NatC type. Besides, At1g24040 acetylate more NatF type substrates than pNaa10p but both proteins shared Met-Tyr- (MY-) as the specific substrate of plastidic NATs (Fig. 3.42). The difference between substrate specificity of At1g24040p and pNaa10p is mainly at the position after acetylated amino acid (Fig. 3.43B). As the second pNATs, At1g24040p substrate specificity had strong correlation with cTP sites (Fig. 3.43C) and we named this protein pNaa20p and its activity pNatB.



Substrates	A	S	T	V	MD	ME	MN	MF	MI	ML	M(M/A/G/S/V)	MQ	MK	MY
No.	20	25	13	1	2	4	4	6	6	10	5	2	2	1

Figure 3.42 At1g24040p substrates identified by LC-MS/MS.

101 substrates of At1g24040p were identified by LC-MS/MS from At1g24040p expressing *E. coli*. Substrate type of NatA and NatC are predominant. At1g24040p and At2g39000p share MY- as the specific substrate of plastidic NATs.



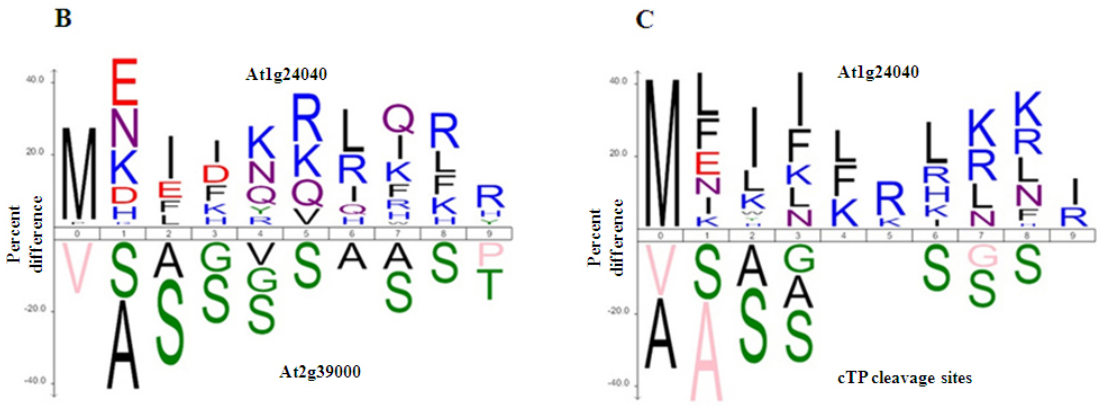


Figure 3.43 Substrate specificity of At1g24040p in comparision with pNaa10p and cTP cleavage sites. (A) WebLogo of At1g24040p substrate specificity (n=101). Bits indicate the sequence conservation at amino acid position, $\log_2 20 = 4.3$ bits (B) IceLogo for the substrate specificity comparision of At1g24040p *vs* pNaa10p (At2g39000). (C) IceLogo for the comparision of At1g24040p substrate specificity *vs*. cTP cleavage site.

3.3.5.3 Nt-acetylation by At2g06025 (pNaa30p)

a. Identification of recombinant At2g06025p in the sample

Similar to pNaa10p and pNaa20p, the presence of At1g24040p in the sample was confirmed by LC-MS/MS with 38% of protein sequence coverage. Based on the emPAI of 6.17 and the total ∑emPAI for the experiment of 2050, the amount of At1g24040p in the lysate was estimated about 0.3%. None of lysine residues was found acetylated (Fig. 3.43 and Table 3.6).

```
  1 MSTISIHRTE TLRITHARSR IFRQRYRRTI PLWKLTINSR SSDTSKKEEL
 51 SVQISIPPQV DQSRPEGLRF DRLQPPEPEF GHEDRFEFGK FVAREAMLDE
101 EYWTAAWLRA ESHWEDRSNE RYVDNYKRKF AEQEFNAIKR RCKGMQGQKC
151 SCIVAVKKEE KHIKRSVIKS VVGTLDLSIR YFLQGETFPG EKVKSQLFCS
201 INQEGSNRYG YIANLCVAKS ARRQGIACNM LRFAVESARL SGVEQVYVHV
251 HKNNSVAQEL YQKTGFKIVE TGKFESLDDD TYLLQYTS
```

Figure 3.43 Identification of At2g06025 by MS.

At2g06025p in sample was identified by MS with 38% amino acid sequence coverage. Matched peptides were highlighted in red.

Table 3.6 Identified sequences of At2g06025 by MS

Start	End	Mascot Score	Expect	Peptide
122	128	36	0.00051	R.YVDNYKR.K
129	140	83	8.30E-09	R.KFAEQEFNAIKR.R
142	165	29	0.0049	R.CKGMQGQKCSCIVAVKKEEKHIKR.S
166	180	81	7.10E-09	R.SVIKSVVGTLDLSIR.Y
181	208	64	1.20E-06	R.YFLQGETFPGEKVKSQLFCSINQEGSNR.Y
209	222	65	5.90E-07	R.YGYIANLCVAKSAR.R
223	232	37	0.00036	R.RQGIACNMLR.F

b. Substrate specificity of At2g06025p (pNaa30p)

Unlike pNaa10p and pNaa20p, very limited substrates were identified in At2g06025p expressing *E. coli*. In total of around 200 quantified N-termini, only 15 N-ternimi with Nt-acetylation level higher than in the control sample. Most of substrates (14) belonged to NatA and NatB type, one substrate was Met-His-, which is a specific substrate of pNATs (Fig.3.44). Nonetheless, At2g06025p displayed N^{α}-acetyltransferase activity, thus we named this protein pNaa30p and its activity pNatC.

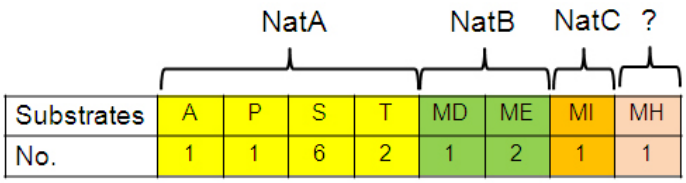


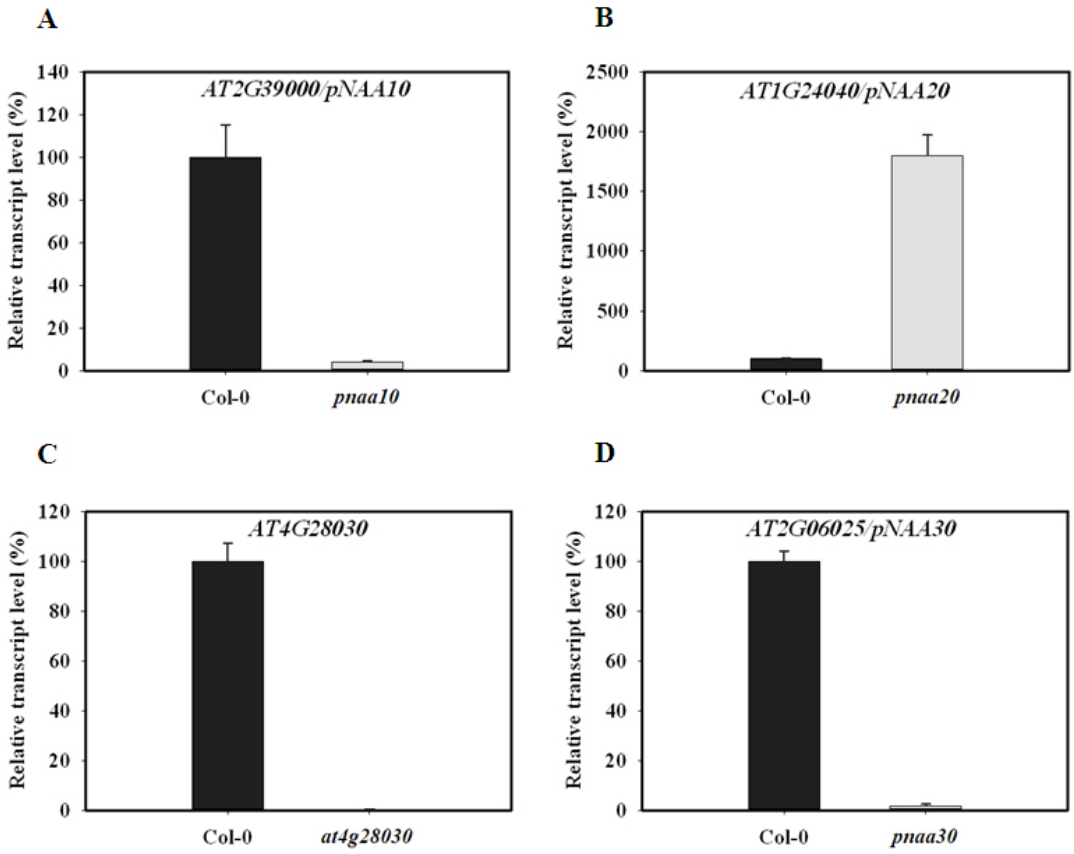
Figure 3.44 At1g24040p substrates identified by LC-MS/MS.

Limited substrates of At06025p were identified by LC-MS/MS from At2g06025p expressing *E. coli*. 14 substrates are NatA and NatB type, 1 substrate is Met-His- (MH), a specific substrate for pNATs.

3.3.6 Identification and characterization of T-DNA insertion mutants of pNATs

To further characterize the significance of plastidic N^{α}-acetyltransferases in *Arabidopsis thaliana*, different T-DNA insertion lines were isolated from Nottingham Arabidopsis Stock Centre (NASC) (Table 2.1 and Fig. S1). The homozygote plants were identified by PCR genotyping (section 2.4.2) with gene specific primers and T-DNA insertion primers (listed in section 2.1.6). In total of seven mutants, six homozygote lines [SALK_072318 *(at2g39000/pnna10),* GK-034E09.02 *(at1g24040/paaa20),* SALK_053123 (*at2g06025 /pnaa30),* WiscDsLox4B08 (*at4g28030),* SALK_032239 (*at1g32070)* and SALK_062388C *(at1g26220)*] were obtained. Unfortunately, all plants from GK-809D05 *(at4g19985)* seeds were wild type, no T-DNA insertion was detected by PCR genotyping (Fig. S2).

To assess the effect of T-DNA insertion on gene expression, qRT-PCR with specific primers (section 2.1.6 and Fig. S1) was applied (section 2.4.8). As shown in Fig. 3.43, the expression of *pNAA10, pNAA30* and *AT1G26220* were severely disrupted. *AT4G28030* and *AT1g32070* were knocked out by T-DNA insertion. However, T-DNA insertion at the 5'-UTR increased the expression of *pNAA20* up to 17-fold than in the wild type.



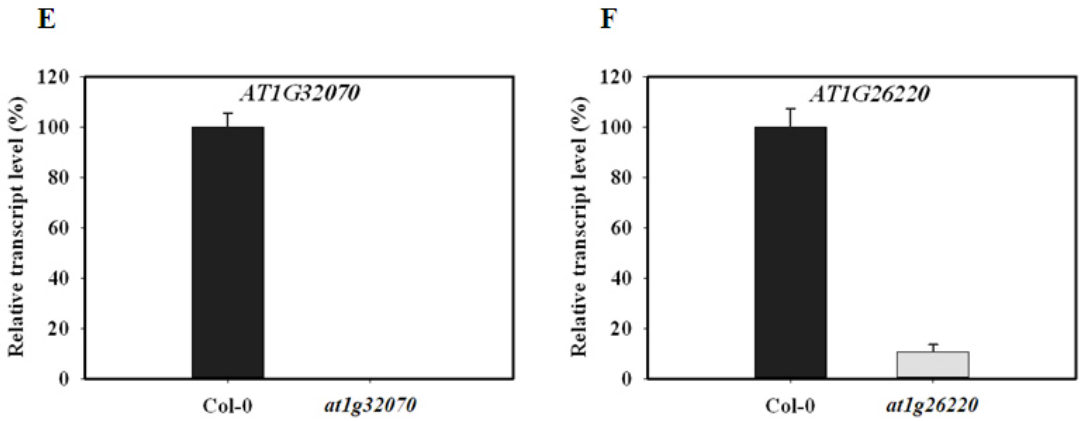
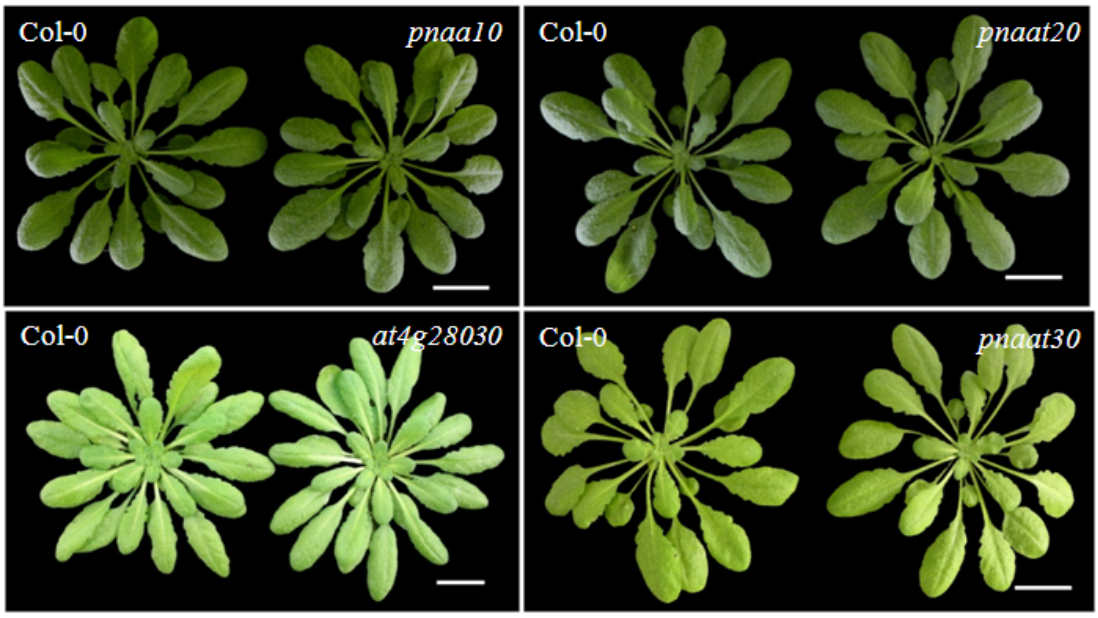


Figure 3.45 qRT-PCR analysis of pNATs T-DNA insertion mutants.

cDNAs of homozygote mutants (*at2g39000/pnaa10, at1g24040/pnaa20, at2g06025/pnaa30, at4g28030, at1g32070, at1g26220*) were used for pRT-PCR analysis. (A, D and F) Expression of *pNAA10, pNAA30* and *AT1G26220* were severely disrupted by T-DNA insertion. (B) T-DNA insertion at the 5'-URT increased the expression of *pNAA20* up to 17-fold. (C and E) *AT4G28030* and *AT1g32070* were knocked out by T-DNA insertion. Error bar indicates standard error (n=3).

Although pNaa10p and pNaa20p are likely to acetylate abundant proteins in the chloroplasts (section 3.3.5), neither functional disruption of *pNAA10* nor over-expression of *pNAA20* showed any effects on the phenotypes under normal growth conditions (Fig. 3.46). Besides, *pnaa30* and other mutants of putative pNATs (*at4g28030, at1g32070, at1g26220*) also displayed the same phenotypes in comparison to the wild type (Fig. 3.46).



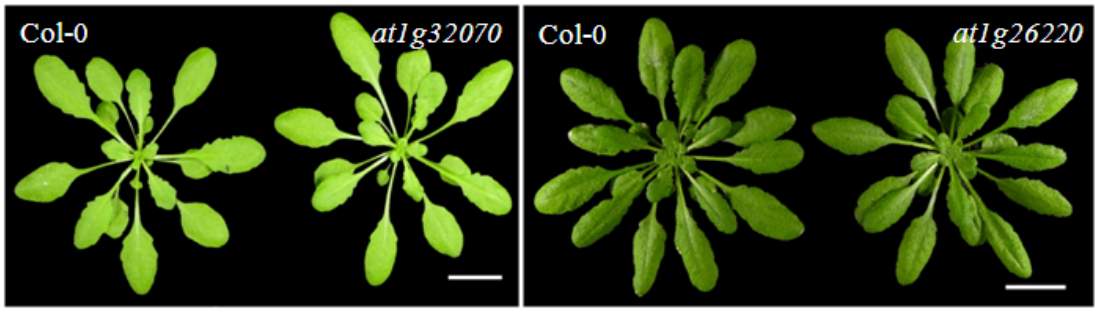


Figure 3.46 Phenotypes of pNATs T-DNA insertion mutants.

Seeds of different mutants were planted in the soil and grown under short day conditions. 7-8 week-old plants were used for documentation. White bar indicates scale of 2 cm.

3.3.7 Expression of plastidic N$^{\alpha}$-acetyltransferase coding genes under drought condition

Very recently, it was reported that Nt-acetylation has potential involvement in the quantitative regulation of the ε-subunit of chloroplast ATP synthase under drought stress (Hoshiyasu et al. 2013). In order to investigate the behaviour of plastidic NAT encoded genes under dought stress, wild type plants were exposed to drought condition (section 2.2.4), cDNA synthesized from total RNA of drought plants was used for qRT-PCR analysis (section 2.4.8).

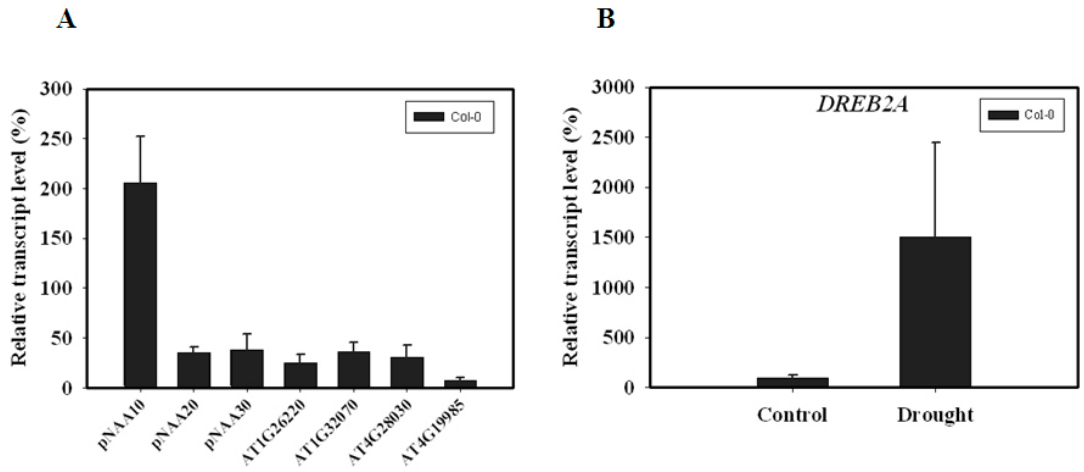


Figure 3.47. Expression of different plastidic NAT-encoded genes under drought tress.

Five week-old Col-0 plants were grown in short day conditions without watering for 2 weeks, wheres the control plants were watered regularly. Gene expression of was analyzed by qRT-PCR. The error bar indicates standard error (n=3). (A) Expression of plastidic NAT-encoded genes. The expression of pNAT-encoded genes were default as 100% in the control condition (B) As the marker for drought stress conditions (Sakuma et al. 2006), the expression of *DREB2A* was increased 15-fold.

The relative transcript level of different plastidic NAT coding genes was shown in the Fig. 3.47A. Under drought stress, the expression of *pNAA10* was increased two folds, whereas the expression of other *pNATs* was down regulated. Especially only seven percents of *AT4G19985* was transcribed. As the marker for drought stress (Sakuma et al., 2006), the expression of *DREB2A* was increased 15-fold than in the control (Fig. 3.47B).

4. DISCUSSION

Since the first discovery more than 50 years ago (Narita, 1958), Nt-acetylation was shown to be one of the most common protein modifications in eukaryotes, occurring on approximately 50-70% of yeast soluble protein and about 80-90% of human protein (Driessen et al., 1985; Flinta et al., 1986; Jornvall, 1975). Nt-acetylation in yeast and in human is thoroughly investigated with the identification of five (NatA-NatE) and six (NatA-NatF) NAT types, respectively. In contrast, the knowledge of Nt-acetylation in plants is vacant for many years. The first Arabidopsis NAT, AtNatC was identified in 2003 (Pesaresi et al., 2003), and very recently three more NATs (NatA, NatB and NatE) were described by Stephan (2011). In this study, we identified two NATs (NatD and NatF) that are still missing in plants and three novel NATs (pNatA, pNatB and pNatC) that are responsible for post-translational Nt-acetylation in chloroplast.

4.1 Protein Nt-acetylation by AtNatD

Naa40p/NatD is unique NAT type that consists of only catalytic Naa40p and acetylates histone H4 and H2A. Homologues of NatD are present in all eukaryotes and its activity is so far characterized in yeast, human and Arabidopsis. NatD is believed to be conserved among species with respect to protein sequence, subcellular localization and substrate specificity.

4.1.1 AtNaa40p localizes to the cytoplasm and to the nucleus

yNaa40p was found to be located in the cytoplasm both in ribosome-bound form and free ribosome bound form (Polevoda et al., 2009b). However, hNaa40p localized not only to the cytoplasm but also to the nucleus. To localize hNa40p, Hole et al. (2011) had separated different subcellular fractions, endogenous hNaa40p was detected both in the cytosolic and nuclear fraction. In agreement with this study, the transient expression of AtNaa40p-EYFP in *Arabidopsis thaliana* showed the present of this protein both in the cytoplasm and in the nucleus (Fig. 3.6).

Cytosolic localization suggests a co-translational acetylation of nascent polypeptides, and therefore Naa40p is physically associated with ribosomes. Indeed, both yNaa40p and hNaa40p were demonstrated to associate with ribosomes (Polevoda et al., 2009b; Hole et

al., 2011). However, only a small amount of Naa40p was found in the ribosome-bound form, mainly Naa40p was free in the cytoplasm or in the nucleus. The presence of larger amount of free Naa40p in the cytoplasm can be explained. Naa40p is synthesized in the cytoplasm but not all of this protein is involved in acetylation of nascent polypeptides at the same time. Since only active Naa40p is associated with ribosomes, most of Naa40p is found free in the cytoplasm.

In contrast to cytosolic localization, the nuclear localization indicates a potential role of AtNaa40p in the nucleus as well as potential function as a lysine acetyltransferase against protein histone. Since hNaa50p displays both N^{α}-acetyltransferase and N^{ε}-acetyltransferase activities (Evjenth et al., 2009), the potential function of Naa40p as a KAT in nucleus is a promising investigation in the future.

4.1.2 N^{α}-acetyltransferase activity of NatD

NatD was shown to acetylate the N-termini of conserved histone H4 and H2A. yNaa40p was discovered in the search for the acetyltransferase responsible for Nt-acetylation of histone H4. Histone H2A was also identified as a substrate of yNaa40p in the same study (Song et al. 2003). Later, hNaa40p was reported to N-terminally acetylate histone H4 and H2A peptides *in vitro* (Hole et al., 2011). In agreement with hNaa40p and yNaa40p, our study showed that AtNaa40p also preferred histone H4 mimicked peptide as the substrate (Fig. 3.3). However, the situation of histone H2A in Arabidopsis is more complicated than in yeast and human. There are 12 different isoforms of histone H2A with various N-termini in Arabidopsis (Table 3.7). Based on the substrate specificity of hNaa40p, only three isoforms of histone H2A (HTA3, HTA5 and HTA9) might be acetylated by AtNaa40p. Since the Nt-acetylation of histone in plants is scarcely known, the role of AtNaa40p in histone acetylation need to be further investigated with respect to both NAT and KAT activity.

Table 4.1 Different isoforms of histone H2A in *Arabidopsis thaliana.*

No.	Locus	Histone H2A isoform	N-termini
1	AT4G27230	HTA2	AGRGK
2	AT1G54690	HTA3	SSGAG
3	AT4G13570	HTA4	IMVEG
4	AT1G08880	HTA5, H2AX	STGAG
5	AT5G59870	HTA6	ESTGK
6	AT5G27670	HTA7	ESSQA
7	AT2G38810	HTA8, H2A.Z	AGKGG
8	AT1G52740	HTA9, H2A.Z	SGKGA
9	AT1G51060	HTA10	AGRGK
10	AT3G54560	HTA11, H2A.Z	AGKGG
11	AT5G02560	HTA12	DSGTK
12	AT3G20670	HTA13	AGRGK

In addition to perhaps acetylating only two selected substrates, H2A and H4, there is difference of points of view about the length of nascent polypeptide required for efficient acetylation by Naa40p. Polevoda et al. (2009b) postulated that yNatD requires a significantly long N-terminal sequence, between 23 and 51 amino acid residues, to efficiently acetylate a proper substrate *in vivo*. In contrast, Hole et al. (2011) concluded that for *in vitro* assay, seven amino acids are clearly enough to ensure the specificity and acetylation by hNaa40p, the difference in requirement for more residues in yeast experiments might implicate differences between species. Although our data approve of Hole's conclusion, we have a different point of view about this disagreement. In our opinion, the difference between yNaa40p and hNaa40p greatly reflects different requirements between an *in vitro* setting with purified enzyme and peptides, and the *in vivo* setting in yeast where yNaa40p is bound to the ribosome. In *in vivo* conditions, enzyme activity is regularly regulated. The interaction of Naa40p with ribosome or with any proteins might alter its allosteric structure which determines substrate specificity. However, in *in vitro* conditions, only chemical properties decide the direction of reaction. That might explain why hNaa40p can *in vitro* acetylate hundreds of peptides but so far only two substrates of NatD have been identified. In approval of our perspective, it was shown that purified hNaa10p and hNatA complex displayed different substrate specificities

when they were exposed with proteome-derived peptide library (Van Damme et al., 2011b). In addition, the difference in amino acid length requirement for *in vitro* and *in vivo* acetylation of Naa40p even suggests an existence of auxiliary subunit which has been failed to identify or the alternative role of ribosome as the auxiliary subunit of NatD.

4.1.3 Histone N-tail and phenotypes of *naa40* mutant

Under the normal growth conditions, both in yeast and in Arabidopsis, the *naa40* knockout mutant did not show any observable phenotypes (Song et al., 2003; section 3.1.5.3). The *ynaa40* mutant grew normally at both rich medium and synthetic medium. A *ynaa40/naa40* diploid underwent normal sporulation and meiosis, leading to four viable haploids per spore (Song et al., 2003). In our research, two *atnaa40* T-DNA insertion mutants that overexpressed and down regulated *ATNAA40* both displayed the normal growth, normal development and ability of producing seeds (Fig. 3.10). However, on the media with additions of 3-amino-1,2,4-triazole, a general inhibitor of transcription, the *naa40-Δ* strains exhibited a minor reduction in growth. In addition, the germination of *atnaa40-1* seeds was reduced when they were grown on the MS medium supplemented with NaCl (3.11). The phenotypes of *naa40* mutants might be the result of either or both of following two reasons: 1) the diminished activity of histones H4 and H2A due to lack of Nt- acetylation; 2) the lack of Nt-acetylation of unidentified proteins is directly or indirectly responsible for these phenotypes.

In order to address the mechanism caused the phenotypes in yeast *naa40-Δ* strains, Polevoda et al. (2009b) have used two other mutants: *hhf2-100* and *hhf2-102*. The *hhf2-100* is the mutant in which three lysine residues K5, K8, and K12 of histone H4 were replaced by arginine. Normally, K5, K8 and K12 are acetylated in histone H4, and these acetylations are prevented by the arginine replacement. The double mutant *hhf2-100 naa40Δ* strain grew much slower on the medium containing benomyl and thiabendazole, whereas the moderate effect was observed on the salt containing media. In contrast, the *hhf2-102* strain in which the H4 N-terminal serine residue was replaced by valine which blocks the Nt-acetylation showed the similar phenotypes as the *naa40Δ* strain in different conditions. Therefore, it was clear that, the lack of N-terminal serine acetylation that further increases the overall positive charge of H4 N-tail caused the growth defect on *naa40Δ* strain, not by a lack of acetylation of any other proteins.

In addition, it was reported that the N-tail of histone H4 is important in DNA damage response (Bilsland and Downs, 2005). Four lysines, at the positions 5,8, 12 and 16, in the N-terminal tail of H4 are reversibly acetylated in vivo in all eukaryotes (Bird et al., 2002). The yeast strain carrying mutations in all H4 N-tail lysines (*hhf1-10*) was markedly hyper-sensitive to both camptothecin (CPT) and methyl methane sulfonate (MMS) but not UV, suggesting the presence of a defect in DNA double-strand break (DSB) repair but not a global defect in all DNA repair processes. Our investigation into the role of H4 N-tail in DNA damage response in plant showed that both *atnaa40-1* and *atnaa40-2* mutants are not sensitive to UV-c in comparison with the wild type (Fig. 3.12). However, the un-sensitiveness of *atnaa40* mutants to UV-c does not exclude the role of Nt-acetylation of H4 in DNA damage response. To further assess this issue, preventing the other acetylation of lysine residues at the H4 N-tail by point mutation has to be considered. Besides, other DNA damage agents such as CTP, MMS or reactive oxygen species might be used for investigation as well.

4.1.4 Substrates of NatD

Histones H4 and H2A were reported to be N-terminally acetylated by yNatD (Song et al., 2003). The *in vitro* experiments strongly suggest histones H4 and H2A as the substrate of hNatD and AtNatD (Hole et al., 2011; section 3.1.3). Besides histone H4 and H2A, so far there is no evidence that demonstrates Naa40p acetylating any others yeast proteins, as judged by 2D gel analysis of the soluble proteins of the wild type and *ynaa40*Δ mutant strains (Song et al., 2003). The phenotypes of *naa40*Δ strains were caused by lack of acetylation of serine at the N-terminal of H4, not by lack of acetylation of other unidentified proteins (Polevoda et al., 2009b). Using the NBD-Cl fluorescent assay, we estimated that the level of free N-terminus proteins in the wild type and *atnaa40* mutants are similar indicating a very few substrates are acetylated by AtNaa40p (Fig. 3.13). Taken together, all experimental data suggested a very limited number of proteins acetylated by NatD. The existence of other unidentified substrates is possible and these proteins might start with methionine or glycine residues according to the *in vitro* substrate specificity of hNaa40p (Hole et al., 2011).

4.2 Protein Nt-acetylation by AtNatF

4.2.1 Is the auxiliary subunit of AtNatF a membrane-anchored protein?

Normally each NAT complex consists of the catalytic subunit and up to two auxiliary subunits, both subunits are required for proper enzyme activity. For instance, AtNatA complex composed of the catalytic subunit AtNaa10p and the auxiliary subunit AtNaa15p. Knockout of *ATNAA10* or *ATNAA15* both resulted in lethality of plants at embryo globular state (Stephan, 2011). NatD is a special type of NAT, only catalytic Naa40p is found, the auxiliary subunit Naa45p either does not exist or so far is unidentified (Polevoda et al. 2009b; Hole et al., 2011). However, Naa40p is associated with ribosome suggesting a possibility of ribosome acting as an auxiliary subunit of NatD. In support of this hypothesis, Naa40p required different length of amino acids at the N-terminus for the efficient acetylation *in vitro* and *in vivo* (Polevoda et al., 2009b; Hole et al. 2011). Taken together, the catalytic subunit of NAT requires interaction with the auxiliary subunit or any binding partners to perform efficient activity and assure substrate specificity.

AtNaa60p localizes to the plasma membrane and to the tonoplast but it might be not an integral membrane protein. High concentration of Na_2CO_3 was able to wash AtNaa60p away from membranes (section 3.2.5) suggesting an interaction between AtNaa60p and a transmembrane protein. This transmembrane protein might be the auxiliary subunit of AtNatF. Futhermore, the potential integral membrane auxiliary subunit might constitute an interesting acetylation model of AtNatF in which the hypothetic auxiliary subunit AtNaa65p anchors catalytic subunit AtNaa60p to the membranes, the active centre of enzyme is facing cytoplasm and post-translationally acetylates cytosolic proteins.

In addition, AtNaa60p differs from other Arabidopsis NATs on the additional about 50 amino acid at the C-tail (Fig. 3.14, section 3.2.1) which is not located in acetyltransferase domain (Fig. 3.17, section 3.2.4). Therefore it is possible that AtNaa60p might interact with its partner or plasma membrane at the C-terminus. However, the fusion with YFP at the C-terminus does not affect on the subcellular localization of AtNaa60p (section 2.3.4) suggesting that the interacting region is not located at the very end of C-tail.

4.2.2 N^{α}-acetyltransferase activity of AtNaa60p

NatA, NatB and NatC are currently considered as three major NATs in eukaryotes with acetylation of broad proteins. NatF, another major NAT of higher eykaryotes, *in vitro* and

in vivo acetylated numerous substrates. Remarkably, hNaa60p prefered substrates with N-termini included Met-Lys-, Met-Ala-, Met-Val- and Met-Met-, which are not reported to be acetylated by any known NATs. Besides, Met-Leu-, Met-Ile- and Met-Phe- starting N-termini, a class of considered NatC substrates, were also acetylated by hNaa60p (Van Damme et al., 2011c). With this discovery, the classical clear-cut classification between NAT substrate classes should be re-evaluated. In agreement to this finding, substrate specificities of pNatA, pNatB and pNatC, three NATs responsible for acetylation in chloroplast, are strongly overlaying (section 3.3.5).

From the current knowledge, substrate specificity of NATs is conserved among species (Polevoda et al., 2009a; Starheim et al., 2009; Hole et al., 2011; Van Damme et al., 2011c; Stephan, 2011). Thus, it was believe that substrate specificity of AtNatF is also conserved from human. Indeed, purified MBP-AtNaa60p was able to acetylate the N-terminus of AKR1 protein, a substrate of hNaa60. However, AtNaa60p not only acetylated methionine residue but also serine residue, a predicted NatA type substrate (section 3.2.3). The acetylation of serine suggested an enigmatic substrate specificity of AtNaa60p which is not only based on hNaa60p. Further investigation using *in vivo* system is necessary to clarify this unrevealed point.

4.2.3 Relation between localization and phenotypes of *atnaa60* mutant

AtNaa60p is believed to acetylate broad substrates but *atnaa60* T-DNA insertion mutant did not showed any phenotypes under the normal growth conditions. The *atnaa60* plants grew normally, reached the same size as the wild type and produced fertilized seeds (Fig. 3.27). The dual localization of AtNaa60p to the plasma membrane and tonoplast suggests a role of this protein in trafficking and osmotic stress response. Therefore, the phenotypes of *atnaa60* might be uncovered under high salt or osmotic tress conditions. Indeed, on the MS media supplemented with NaCl, the germination of *atnaa60* seeds were significant slower than the wild type (Fig. 3.28). Surprisingly, *atnaa60* seeds germinated at the same level as the wild type on the media supplemented with 165 mM mannitol (Fig. 3.28) indicating the specific effect of salt on *atnaa60* mutant. Thus, AtNaa60p might be involved in Na^+ transport. Our further investigation revealed that *atnaa60* mutant was only sensitive to salt stress during the germination stage but not seedling development (Fig 3.29). In addition, *ATNAA60* was also expressed the same level in salt stressed and non-salt stressed seedlings (Fig. 3.30). Germination is a special stage of plant development when seeds imbibe a lot of

water for swilling, breaking the seed coat and breaking down the stored nutrition. During this stage, seeds are very sensitive to high salt concentration or osmotic causing factors. When plants grow big enough, they develop mechanisms to help them resist these conditions. Taken together, high NaCl concentration had a slight effect on *atnaa60* mutant so that *atnaa60* seeds were struggling to germinate but no phenotypes were able to observe during seedling development. Additionally, this minor phenotype seems to be related to AtNaa60p localization, not by lacking of acetylation of any proteins.

4.2.4 NatF N-terminally acetylates numerous proteins

hNaa60p acetylated a plenty of proteins, contributing about 10 percents to overall protein acetylation when it was over-expressed in yeast (Van Damme et al., 2011c). Thus, it was believed that Naa60p acetylates the same amount of proteins in higher eukaryotes. In agreement with this judgment, a large amount the free N-terminal proteins which are supposed to be acetylated by AtNaa60p were found in *atnaa60* mutant using NBD-Cl fluorescent assay (Fig. 3.31). Together with three other NATs (NatA, NatB and NatC), NatF is currently considered as the major NAT in higher eukaruotes.

AtNaa60p potentially has broad substrate specificity but which type of protein is acetylated by AtNaa60p is still scarcely known. Our *in vitro* acetylation assay revealed that proteins with N-termini Met-Val- and Ser- are the potential AtNaa60 substrates. The unusual localization, together with the disagreement between AtNaa60p substrates *in vitro* and hNaa60 substrate data set, prove the unclarity of AtNaa60p substrates which cannot just based on substrate specificity of hNaa60 to judge.

4.3 Protein Nt-acetylation by AtpNATs

4.3.1 Bacterial proteins – an unrestricted source for Nt-acetylation

Since the first NAT was identified in 1989 (Mullen et al. 1989), plentiful methods and techniques have been used and developed to identify NAT substrates and substrate specificity. In early studies two dimensional (2D) protein gels was used to compare maps of acetylated and non-acetylated proteins. By comparing the proteomes of wild type and the mutant deficient for a specific NAT, substrates were revealed by the charge shifts due to the absence or presence of positively charged α-amino groups. However, because of drawbacks of 2D polyacrylamide gel techniques, only well soluble and highly abundant

proteins were detected. For instance, NatB was estimated to acetylate about 1000 proteins but only 4 substrates were identified by comparing 2D gel maps of wild type and *atnaa20* proteomes (Stephan 2011). Later, gel-free proteomics techniques were applied to cover more comprehensive proteome (Gevaert et al. 2007). In this, proteomes are digested into peptides which are more soluble than their precursor proteins and can be readily identified and quantified by mass spectrometry. A development of gel-free proteomics techniques was COFRADIC (combined fractional diagonal chromatography) that enriches N-terminal peptides (Van Damme et al., 2009). However, this method requires a comparison of proteomes of wild type and the mutant deficient for a specific NAT. Thus, in case of redundancy of NATs like pNATs, when the knockout of one can be complemented by the others, the lack of substrate identification is inevitable.

In parallel with *in vivo* approaches, synthetic peptides and proteome-derived peptide library (Van Damme et al., 2011b) were used to identify substrate specificity of NATs *in vitro*. Despite the fact of simplicity, *in vitro* methods have intrinsic drawbacks which are unavoidable. First, *in vitro* assay requires a purification of the potential NAT which is not possible in all cases. Second, due to lack of tools and high cost, not too many synthetic peptides can be used in *in vitro* assay, therefore these peptides might be not acetylated although the investigated protein is a true NAT. Proteome-derived peptide library overcomes this disadvantage of synthetic peptides and becomes an efficient tool for substrate specificity identification. However, results from *in vitro* and *in vivo* system are not always in agreement due to lack of regulation or lack of auxiliary subunit which is also determine specific substrates of NAT. For instance, Naa40p required different length of N-terminus for efficient acetylation of histone H4 *in vitro* and *in vivo* (Van Damme et al., 2009b; Hole et al., 2011).

In this study, we used for the first time the presence of numerous free N-termini of *E. coli* proteome to identify substrate specificity of pNATs. The recombinant NATs are able to use these proteins as their substrates. The acetylated N-termini were subsequently identified by LC-MS/MS. This method, on one hand remedies the disadvantages of *in vivo* assay, on the other hand can avoid the complement caused by the redundancy of NATs. One drawback of our approach is the differences in regulation between eukaryote and prokaryote cells might have effect on acetylation. In fact, substrate specificity of three pNATs (pNatA, pNatB and pNatC) was successfully identified using this novel approach.

This method would be useful in the future for *in vivo* identification of substrate specificity of other NATs.

4.3.2 Nt-acetylation in chloroplast and plastidic NATs

In yeast and human, so far acetylation of protein is known to take place only in cytoplasm. Five NAT complexes in yeast and six NAT complexes in human are responsible for the acetylation of these proteins (Poveloda et al., 2009a; Van Damme et al., 2011c). With the identification of human NatF, it is likely that all human NATs acting co-translationally have been identified since all classes of substrates found to be acetylated are covered (Van Damme et al., 2011c). In plants, all homologous NATs from yeast and human have been identidied and these enzymes also acetylate cytosolic proteins (Pesaresi et al., 2003; Stephan, 2011; section 3.1 and 3.2).

However, plant, a photoautotrophic organism possesses chloroplast that neither yeast nor human has. The acetylation of protein is reported to also occur in chloroplast (Kleffmann et al., 2007; Zybailov et al., 2008; Bienvenut et al., 2012). Up to now, 4 chloroplast-encoded proteins and more than 220 nuclear-encoded proteins are known to be acetylated in Arabidopsis chloroplast (Zybailov et al., 2008; Bienvenut et al., 2012). In this study, we have identified three plastidic NATs (pNatA, pNatB and pNatC - section 3.3). Substrate specificity of pNatA and pNatB (Fig. 3.39 and 3.41) shows strong correlation to the experimentally characterized acetylated chloroplastic proteins (Fig. 4.1). Ala, Ser, Val and Thr residues are predominantly acetylated in both caces, only minor number of proteins starting with Ile, Asn and Glu are not acetylated by pNatA and pNatB in *E. coli*. The differences between chloroplastic proteome and *E. coli* proteome or the existence of other unidentified plastidic NATs could explain for these minor uncorrelations. pNatC displayed similar substrate specificity as pNatA and pNatB but for some reason only acetylated very limited number of proteins (section 3.3.5.3). The acetylation of protein by pNatC has to be further investigated.

The general functions of Nt-acetylation of protein is not well understood, the biological roles of acetylation of chloroplastic protein is even scarcer. So far most of the researches mainly aimed to characterize the N-terminal acetylation of chloroplastic proteins, and Nt-acetylation of protein was revealed as the common post modification in chloroplast (Zybailov et al., 2008; Bienvenut et al., 2012). Very recently, Nt-acetylation of chloroplastic protein was reported to have involvement in the quantitative regulation of the

ε-subunit of chloroplast ATP synthase under drought stress (Hoshiyasu et al., 2013). The role of chloroplastic protein acetylation in drought stress is further investigated and dissccussed in this study (sections 3.3.7 and 4.3.6). Based on the acetylated proteins in chloroplast, Nt-acetylation seems to be related to photosysthetic processes and oxidative stress since several major components of the photosynthetic apparatus, including large subunit of ribulose bisphosphate carboxylase/oxygenase, and photosystem II reaction centrer proteins D1 and D2 were found N-terminally acetylated (Zybailov et al., 2008). In addition, following the plant N-end rule pathway (Graciet and Wellmer, 2010) Nt-acetylation in general might also contribute to the stabilization of proteins since it modifies the N-terminus of protein which cannot further ubiquitinated and subsequently degraded following ubiquitin pathway.

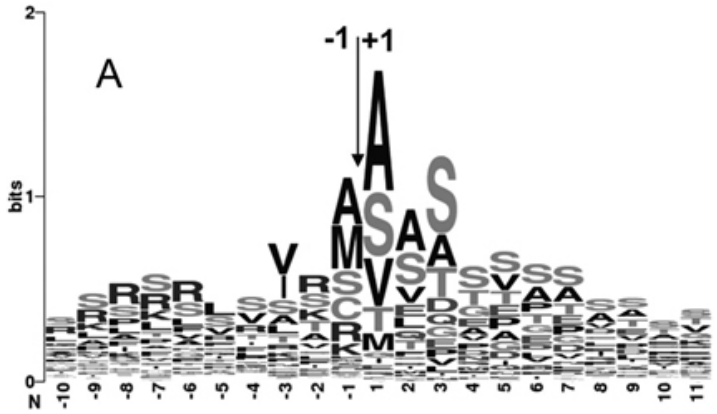


Figure 4.1 Sequenlogo of cTP cleavage sites related to the experimentally characterized acetylated chloroplastic proteins (n=170) (Bienvenut et al., 2012). Bits indicate the sequence conservation at amino acid position, $\log_2 20 = 4.3$ bits

4.3.3 N^{α}-acetyltransferase activity of pNATs

Six NATs, NatA-NatF, are responsible for Nt-acetylation of proteins in cytoplasm of eukaryotes. Each complex differs from the other by specific subunits and substrate specificity. For instance, NatA acetylates proteins with Ser-, Ala-, Thr-, Val-, Gly-, Cys-, and Pro- at the N-termini after removal of Met; NatB, NatC and NatF acetylate Met followed by different amino acid residues (Van Damme et al., 2011c). In contrast to clear-cut substrate specificity of cytosolic NATs, pNATs share their substrates with each other. pNatA and pNatB acetylate broad proteins which cover substrates of all known NATs (section 3.3.5.1 and 3.3.5.2). The situation of pNATs is similar to SsArd1p, a homologue of Naa10p in *Sulfolobus solfataricus.* Besides displaying NatA activity (acetylation of Ala

and Ser residues), SsArd1p also acetylates NatB and NatC substrates (Met-Glu and Met-Leu) (Mackay et al. 2007). The similarities in substrate specificity between pNATs and an archaeal NAT might prove an ancient origin of these proteins.

Unlike other pNATs, pNatC only acetylates a small subset of proteins (section 3.3.5.3) suggesting a high selection of substrates. However, non-fully functional recombinant pNaa30p might be the reason why only a few proteins are acetylated. pNaa30p was surprisingly predicted by TargetP to localize both to the mitochondria and chloroplast with a 111 amino acids transit peptide although it only located in chloroplast (Fig. 3.32). The 111 amino acids transit peptide was cut off before cloning to avoid any effects on enzyme activity. Thus, false prediction of TargetP might result in non-fully functioning recombinant pNaa30p by lack of important domains or containing unnecessary domains. Producing other constructs encoding recombinant pNaa30p with longer or shorter sequence can confirm substrate specificity of pNatC.

In addition, both chloroplast encoded proteins and nuclear encoded proteins are N-terminally acetylated demonstrating different scenarios of acetylation in chloroplast. Mitochondrium and chloroplast are special organelles that have their own ribosomes and translate its own proteins. Thus, chloroplast-encoded proteins might undergo co-translational or post-translational acetylation processes. In co-translational acetylation, the potential auxiliary subunits of pNATs might be required. However, nuclear-encoded proteins are only post-translationally acetylated after the removal of transit peptide.

4.3.4 Acetylation of lysine residues is necessary for efficient function of pNaa10p

Four lysine residues (K217, K243, K254 and K256) were found acetylated when pNaa10p was expressed in *E. coli* (section 3.3.5.1). Acetylation of these lysine residues could be the result of pNaa10p autoacetylation. Indeed, the *in vitro* autoacetylation assay showed that purified recombinant pNaa10p is able to acetylated lysine residues of its own (Fig. 3.40). Since pNaa10p can acetylate its internal lysine residues, the ability of this protein acting as lysine acetyltransferase towards other chloroplastic proteins is possible. However, because of the acetylation of internal lysines is quite common in bacteria (Soppa, 2010), bacterial lysine acetyltransferase might also be able to acetylate pNaa10p lysine residues as well.

Noticeably, two of lysine residues (K217 and K243) are located in the acetyltransferase domain (Fig. 4.2), the acetylation of these residues might be necessary for the efficient

function of pNaa10p. This could explain why the signal of pNaa10p was slightly changed in the first 30 minutes but dramatically increased after this time point when two lysines at the acetyltransferase domain were fully acetylated (Fig. 3.40).

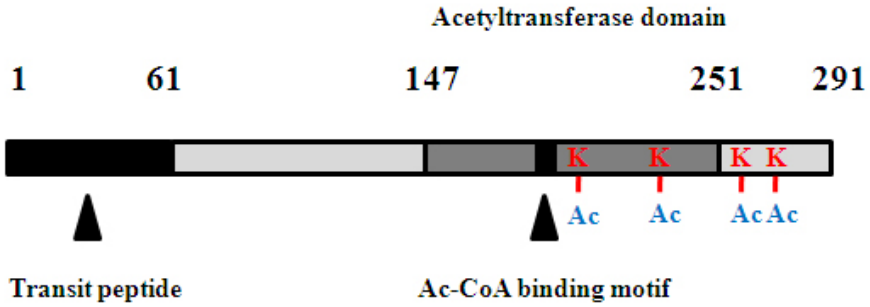


Figure 4.2 Schemetic model of pNaa10p.

pNaa10p is a protein of 291 amino acids with platidic transit peptide from 1-61 amino acid. The acetyltransferase domain locates between amino acids 147-251. Four lysine residues K217, K243, K254 and K265 are acetylated by itself. Transit peptide cleavage site was predicted by TargetP and SMART was used to predict acetyltransferase domain.

4.3.5 Are more than three NATs responsible for acetylation in chloroplast?

As mentioned above, the subcellular localization of putative pNATs revealed seven candidates that might be responsible for the acetylation in chloroplast. Three candidates (At2g39000, At1g24040 and At2g06025), which contain classical Ac-CoA binding motif (RxxGxG/A) were demonstrated to be the genuine NATs and named pNaa10p, pNaaa20p and pNaa30p, respectively.

Three other candidates (At4g19985, At1g26220 and At1g32070) possess modified Ac-CoA binding motif (QxxGxxG/A) where R (arginine) at the first position is replaced by Q (glutamine). Although the modified Ac-CoA binding motif presents only in AtNaa40p, the existence of this motif strongly indicates a potential acetyltransferase activity. Noticeably, these three candidates belong to a small subgroup within group of pNaa20p (Fig 3.35) therefore they might represent for a different type of pNAT which only acetylates a specific class of substrate of its own.

Another plastidic candidate, At4g28030, in which the first amino acid in Ac-CoA binding motif (R or Q) is replaced by histidine (H), an amino acid belongs to positively charged

group like arginine. Since histidine is similar to arginine, it might be able to fulfill the function of arginine in Ac-CoA binding motif. For that reason, At4g28030 is possibly a true NAT.

In conclusion, with four more potential pNATs, chloroplast possesses at least seven different NATs for acetylation of its proteins. The mystery is why an organelle like chloroplast possesses such a high number of NATs, not mention one can cover function of all the others including cytosolic NATs. Furthermore, these plastidic NAT might also responsible for acetylation of internal lysine of chloroplastic proteins.

4.3.6 Role of chloroplastic protein acetylation in drought stress response

Protein Nt-acetylation is one of the most common and abundant protein modifications that affect a large number of proteins. However, the exact biological role has remained enigmatic for majority of affected proteins, and only for a small number of proteins, N-terminal acetyation was linked to various features of protein such as localization, stability and interaction (Hollebeke et al., 2012). Very recently, it was reported that Nt-acetylation has potential involvement in the quantitative regulation of the ε-subunit of chloroplast ATP synthase under drought stress (Hoshiyasu et al., 2013). The ε-subunit of chloroplast ATP synthase presents in two different isoforms, N-teminal acetylated isoform and non-acetylated isoform. Under drought stress, the protein level of non-acetylated isoform preferentially decreased, whereas the abundance of the acetylated isoform remained unchaged.

Since the ε-subunit of chloroplast ATP synthase was post-translationally acetylated in chloroplast and behaved differently from its non-acetylated isoform under drought stress (Hoshiyasu et al., 2013), the regulation of enzymes responsible for this modification might also have changed. Our investigation showed that under dought stress the expression of pNAA10 was increased two-fold, whereas the expression of other pNATs was down regulated, specially only seven percents of AT4G19985 was transcribed (Fig. 4.3). The alteration in regulation of pNATs supports the proposal of Hoshiyasu et al. (2013) about the connection between the Nt-acetylation and regulation of quantitative abundance of acetyled chloroplastic proteins under drought stress. Moreover, the ε-subunit of chloroplast ATP synthase might be one of the substrates of pNatA.

In addition, in chloroplasts, several major components of the photosynthetic apparatus, including large subunit of ribulose bisphosphate carboxylase/oxygenase, and photosystem II reaction centrer proteins D1 and D2 were found N-terminally acetylated (Zybailov et al., 2008) suggesting the involvement of Nt-acetylation in photosysthetic processes and oxidative stress. However, the current knowledge of the occurrence and function Nt-acetylation in photosynthetic apparatus is scarce, requiring further research to elucidate its biological roles.

4.4 Nt-acetylation in *Arabidopsis thaliana*, are there more NATs?

In human, with the characterization of NatF, it is likely that all human NATs acting co-translationally have been identified since all classes of substrates found to be acetylated are covered. However, the circumstances are not the same in plants. In the searches for AtNaa40p, AtNaa60p and putative pNATs, three other putative NATs (At1g03650, At3g02980 and At4g72030) with significant amino acid sequence identity/similarity to known NATs were found.

Remarkably, *AT3G02980* might be the result of *ATNAA60* duplication since they share 76%/89% amino acid sequence identity/similarity (Fig. 4.3A). *AT3G02980* encodes MEIOTIC CONTROL OF CROSSOVERS1 (MCC1), a protein required in meiosis for normal chiasma number and distribution and for chromosome segregation. Interestingly, the role of MCC1 in chromosome segregation is the exact function of Drosophila Naa60p (dNaa60p). Knockdown of dNAA60 resulted in chromosomal segregation defects during anaphase, (Van Damme et al., 2011c). Therefore, it is possible that *ATNAA60* and *AT3G02980*, results of duplication from one original gene, during evolution were selected to have separated functions which are all undertaken by Naa60p in animals and human. The identity of almost entire protein sequence of At3g02980p and AtNaa60p including Ac-CoA binding motif (RxxGxG/A) (Fig. 4.3A) clearly suggests At3g02980 as a true NAT and it might even share part of substrates with AtNaa60p.

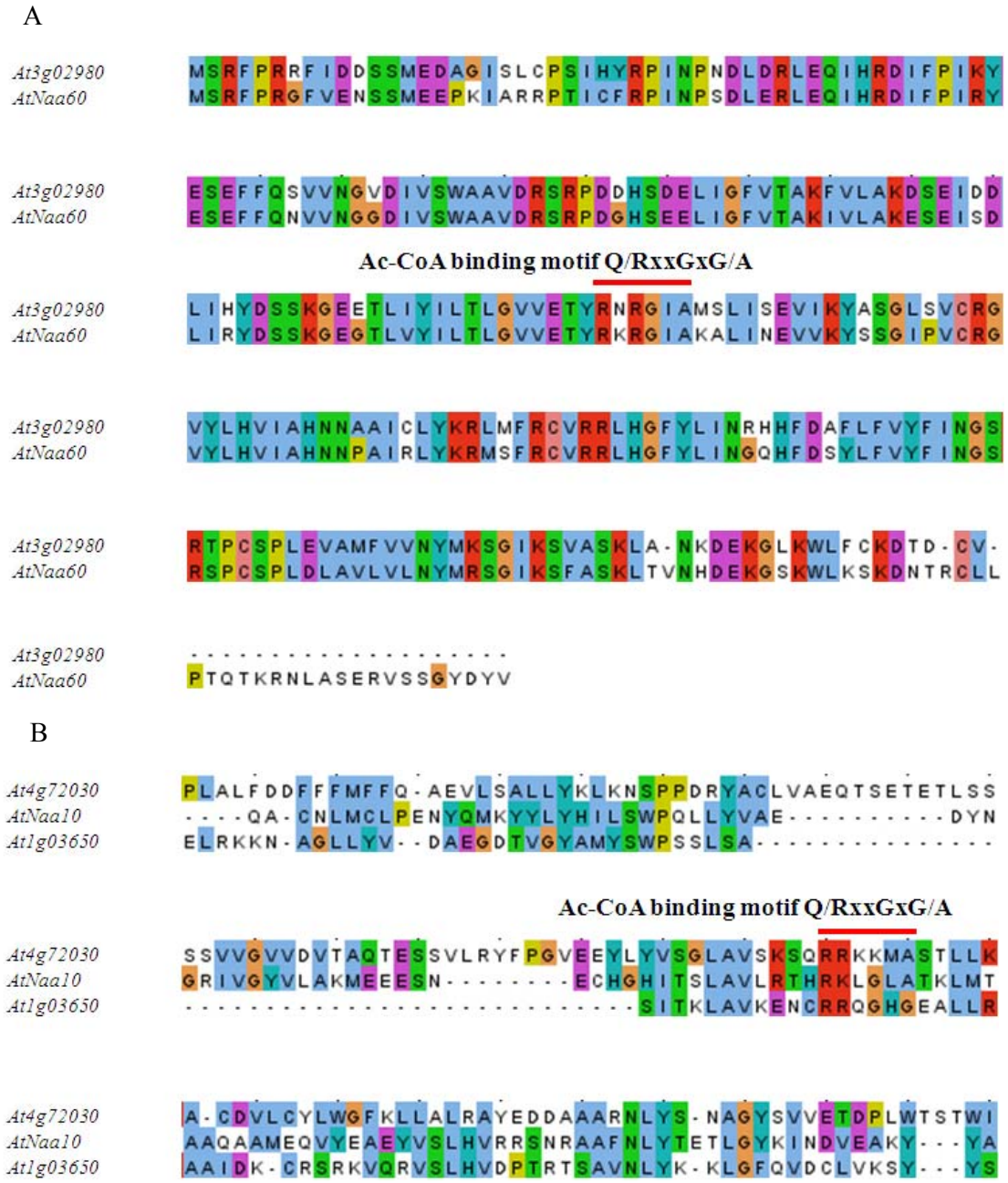


Figure 4.3 Amino acid sequence alignments of other putative NATs.

(A) Amino acid sequence alignment of At3g02980 and AtNaa60p. (B) Amino acid alignment of At4g72030, At1g03650 and AtNatA, only amino acids between 100-250 were used. Clutal Omega (Goujon et al., 20010; Sievers et al., 2011) was used to align all amino acid sequences. The conserved Ac-CoA binding motif Q/RxxGxG/A, where x represents any amino acids is marked with the red line above. The red background indicates basic residues, purple indicates acid residues, orange indicates proline, blue indicates hydrophobic residues, green indicates polar residues, and turquoise indicates histidine and tyrosine residues. Amino acids were highlighted using Jalview (Waterhouse et al., 2009).

In two other putative NATs, only At1g03650 is likely to be a genuine NAT since it shares 34%/48% amino acid sequence identity/similarity with AtNaa10p and contains the classically conserved Ac-CoA binding motif (RxxGxG) (Fig. 4.3B). At4g72030, as mentioned in sections 3.3.1 and 3.3.2, was predicted to localize both in chloroplasts and mitochondria. However, the subcellular localization in Arabidopsis protoplasts by fusion with EYFP showed that At4g72030 possibly localizes to the mitochondria (Fig. 3.32). In the conserved Ac-CoA binding motif of At4g72030 the glutamine, an uncharged amino acid is replaced by lysine, a positively charged amino acid. The replacement of an uncharged residue by a positively charged residue could result in unfunctional motif which leads to the inactivation of N^{α}-acetyltransferase activity. Nonetheless, At4g72030 still need to be further investigated since it might display the differences of mitochondrial NAT.

4.5 The overall picture of protein N-terminal acetylation in plants

Nt-acetylation is one of the most common modifications of protein and is shown to be conserved among species with respect to protein sequence, subcellular localization and substrate specificity. However, besides the conservation, the N^{α}-acetyltransferases also reflect the unique characteristics of particular organism during evolution. In yeast and human, five conserved NAT complexes (NatA-NatE) are responsible for the acetylation of cytosolic proteins at the translational state (Polevoda et al., 2009a). Human and higher eukaryotes additionally express NatF which displays a different type of substrate specificity (Van Damme et al., 2011c). The overall picture of protein N-terminal acetylation in plants is summarized in the Fig. 4.4. In plants, six homologous NATs from yeast and human are found (Pesaresi et al., 2003; Stephan, 2011 and this study). From present data, five of these (NatA-NatE) seems to conserve among yeast, human and plants. Although protein sequence of AtNaa60p is conserved from human, its function is adapted for particular responsibility in plants since it localizes to the plasma membrane and tonoplast. Besides six identified NATs (NatA-NatF), in Arabidopsis two other putative NATs (At1g03650 and At3g02980) might contribute to the acetylation of cytosolic proteins. Most noticeably, plants possess plastidic NATs to undertake the acetylation of protein in chloroplast which neither yeast nor human has. Sofar, three pNATs (pNatA, pNatB and pNatC) have been identified (this study) and these enzymes reflect the differences in substrate specificity between cytosolic NATs and pNATs (section 4.3.3). Furthermore, other putative NATs (At1g26220, At1g32070, At4g19985, At4g28030 and

At4g72030) might be also responsible for the acetylation of proteins in chlotoplasts and mitochondria.

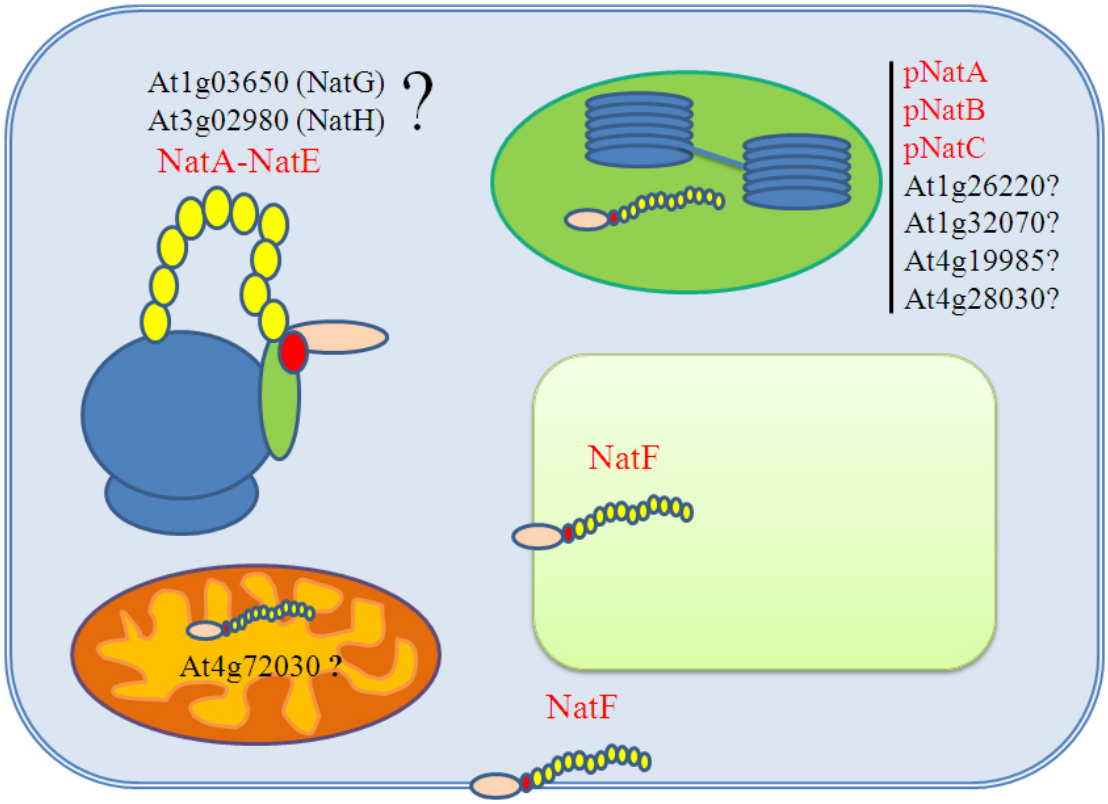


Figure 4.4 Overall picture of protein N-terminal acetylation in plants. The identified NATs are marked in red. The putative NATs are marked in black followed by question mark.

REFERENCES

Aivaliotis M, Gevaert K, Falb M, Tebbe A, Konstantinidis K, Bisle B, Klein C, Martens L, Staes A, Timmerman E, Van Damme J, Siedler F, Pfeiffer F, Vandekerckhove J, Oesterhelt D (2007) Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*. Journal of proteome research **6**: 2195-2204.

Ametzazurra A, Gazquez C, Lasa M, Larrea E, Prieto J, Aldabe R (2009) Characterization of the human N-alpha-terminal acetyltransferase B enzymatic complex. BMC proceedings **4**: 1753-6561.

Andoh T, Hirata Y, Kikuchi A (2000) Yeast glycogen synthase kinase 3 is involved in protein degradation in cooperation with Bul1, Bul2, and Rsp5. Molecular and cellular biology **20**: 6712-6720.

Arnesen T (2011) Towards a functional understanding of protein N-terminal acetylation. PLoS biology **9**: e1001074.

Arnesen T, Anderson D, Baldersheim C, Lanotte M, Varhaug JE, Lillehaug JR (2005) Identification and characterization of the human ARD1-NATH protein acetyltransferase complex. The Biochemical journal **386**: 433-443.

Arnesen T, Betts MJ, Pendino F, Liberles DA, Anderson D, Caro J, Kong X, Varhaug JE, Lillehaug JR (2006) Characterization of hARD2, a processed hARD1 gene duplicate, encoding a human protein N-alpha-acetyltransferase. BMC biochemistry **7**: 13.

Arnesen T, Gromyko D, Kagabo D, Betts MJ, Starheim KK, Varhaug JE, Anderson D, Lillehaug JR (2009) A novel human NatA Nalpha-terminal acetyltransferase complex: hNaa16p-hNaa10p (hNat2-hArd1). BMC biochemistry **10**: 15.

Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, Coker C, Krauskopf A, Kupiec M, McEachern MJ (2004) A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. Proceedings of the National Academy of Sciences of the United States of America **101**: 8658-8663.

Auffret AD, Williams DH (1978) Identification of the blocked N-terminus of an alcohol dehydrogenase from *Drosophila melanogaster* N-11. FEBS letters **90**: 324-326.

Bernal-Perez LF, Prokai L, Ryu Y (2012) Selective N-terminal fluorescent labeling of proteins using 4-chloro-7-nitrobenzofurazan: a method to distinguish protein N-terminal acetylation. Analytical biochemistry **428**: 13-15.

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

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

Bilsland E, Downs JA (2005) Tails of histones in DNA double-strand break repair. Mutagenesis **20**: 153-163.

Bird AW, Yu DY, Pray-Grant MG, Qiu Q, Harmon KE, Megee PC, Grant PA, Smith MM, Christman MF (2002) Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. Nature **419**: 411-415.

Bischof S, (2010) Chloroplast proteome analysis: new insight into intracellular trafficking. Dissertation.

Bischof S, Baerenfaller K, Wildhaber T, Troesch R, Vidi PA, Roschitzki B, Hirsch-Hoffmann M, Hennig L, Kessler F, Gruissem W, Baginsky S (2011) Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. The Plant cell **23**: 3911-3928.

Bonisch C, Hake SB (2012) Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research **40**: 10719-10741.

Brandizzi F, Snapp EL, Roberts AG, Lippincott-Schwartz J, Hawes C (2002) Membrane protein transport between the endoplasmic reticulum and the Golgi in tobacco leaves is energy dependent but cytoskeleton independent: evidence from selective photobleaching. The Plant cell **14**: 1293-1309.

Brown BA, Cloix C, Jiang GH, Kaiserli E, Herzyk P, Kliebenstein DJ, Jenkins GI (2005) A UV-B-specific signaling component orchestrates plant UV protection.

Proceedings of the National Academy of Sciences of the United States of America **102**: 18225-18230.

Brummel MC, Sanborn BM, Stegink LD (1971) Chicken heart H4 lactate dehydrogenase: N-terminal and C-terminal residues. Archives of biochemistry and biophysics **143**: 330-335.

Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods **6(11)**: 786-7. doi: 10.1038/nmeth1109-786.

Coleman MD, Pahal KK, Gardiner JM (1996) The effect of acetylation and deacetylation on the disposition of dapsone and monoacetyl dapsone hydroxylamines in human erythrocytes in-vitro. J Pharm Pharmacol **48**: 401-406.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res **14**: 1188-1190.

Driessen HP, de Jong WW, Tesser GI, Bloemendal H (1985) The mechanism of N-terminal acetylation of proteins. CRC critical reviews in biochemistry **18**: 281-325.

Dutnall RN, Tafrov ST, Sternglanz R, Ramakrishnan V (1998) Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily. Cell **94**: 427-438.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology **300**: 1005-1016.

Evjenth R, Hole K, Karlsen OA, Ziegler M, Arnesen T, Lillehaug JR (2009) Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity. The Journal of biological chemistry **284**: 31122-31129.

Falb M, Aivaliotis M, Garcia-Rizo C, Bisle B, Tebbe A, Klein C, Konstantinidis K, Siedler F, Pfeiffer F, Oesterhelt D (2006) Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey. Journal of molecular biology **362**: 915-924.

Fang H, Zhang X, Shen L, Si X, Ren Y, Dai H, Li S, Zhou C, Chen H (2009) RimJ is responsible for N(alpha)-acetylation of thymosin alpha1 in *Escherichia coli*. Applied microbiology and biotechnology **84**: 99-104.

Flinta C, Persson B, Jornvall H, von Heijne G (1986) Sequence determinants of cytosolic N-terminal protein processing. European journal of biochemistry / FEBS **154**: 193-196.

Fluge O, Bruland O, Akslen LA, Varhaug JE, Lillehaug JR (2002) NATH, a novel gene overexpressed in papillary thyroid carcinomas. Oncogene **21**: 5056-5068.

Forte GM, Pool MR, Stirling CJ (2011) N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. PLoS biology **9**: e1001073.

Gautschi M, Just S, Mun A, Ross S, Rucknagel P, Dubaquie Y, Ehrenhofer-Murray A, Rospert S (2003) The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. Molecular and cellular biology **23**: 7403-7414.

Gevaert K, Impens F, Van Damme P, Ghesquiere B, Hanoulle X, Vandekerckhove J (2007) Applications of diagonal chromatography for proteome-wide characterization of protein modifications and activity-based analyses. The FEBS journal **274**: 6277-6289.

Gordiyenko Y, Deroo S, Zhou M, Videler H, Robinson CV (2008) Acetylation of L12 increases interactions in the *Escherichia coli* ribosomal stalk complex. Journal of molecular biology **380**: 404-414.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. Jul; **38** (Web Server issue):W695-9. doi: 10.1093/nar/gkq313. Epub 2010 May 3.

Graciet E and Wellner F (2010) The plant N-end rule pathway: structure and functions. Trends Plant Sci. **15, (8)**: 447–453.

Gromyko D, Arnesen T, Ryningen A, Varhaug JE, Lillehaug JR (2010) Depletion of the human Nalpha-terminal acetyltransferase A induces p53-dependent apoptosis and p53-independent growth inhibition. International journal of cancer. Journal international du cancer **127**: 2777-2789.

Hatakeyama T (1990) Amino acid sequences of the ribosomal proteins HL30 and HmaL5 from the archaebacterium *Halobacterium marismortui*. Biochimica et biophysica acta **6**: 343-347.

Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the Arabidopsis Subcellular Database. Nucleic acids research **35**: 28.

Helsens K, Van Damme P, Degroeve S, Martens L, Arnesen T, Vandekerckhove J, Gevaert K (2011) Bioinformatics analysis of a Saccharomyces cerevisiae N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation. Journal of proteome research **10**: 3578-3589.

Hershko A, Heller H, Eytan E, Kaklij G, Rose IA (1984) Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. Proceedings of the National Academy of Sciences of the United States of America **81**: 7021-7025.

Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. The EMBO journal **16**: 2054-2060.

Hofmann I, Munro S (2006) An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility. Journal of cell science **119**: 1494-1503.

Hole K, Van Damme P, Dalva M, Aksnes H, Glomnes N, Varhaug JE, Lillehaug JR, Gevaert K, Arnesen T (2011) The human N-alpha-acetyltransferase 40 (hNaa40p/hNatD) is conserved from yeast and N-terminally acetylates histones H2A and H4. PloS one **6**: e24713.

Hollebeke J, Van Damme P, Gevaert K (2012) N-terminal acetylation and other functions of Nalpha-acetyltransferases. Biological chemistry **393**: 291-298.

Hoshiyasu S, Kohzuma K, Yoshida K, Fujiwara M, Fukao Y, Yokota A, Akashi K (2013) Potential involvement of N-terminal acetylation in the quantitative regulation of the epsilon subunit of chloroplast ATP synthase under drought stress. Bioscience, biotechnology, and biochemistry **77**: 998-1007.

Hou F, Chu CW, Kong X, Yokomori K, Zou H (2007) The acetyltransferase activity of San stabilizes the mitotic cohesin at the centromeres in a shugoshin-independent manner. The Journal of cell biology **177**: 587-597.

Hwang CS, Shemorry A, Varshavsky A (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. Science **327**: 973-977.

Jeong JW, Bae MK, Ahn MY, Kim SH, Sohn TK, Bae MH, Yoo MA, Song EJ, Lee KJ, Kim KW (2002) Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation. Cell **111**: 709-720.

Johnson LM, Kayne PS, Kahn ES, Grunstein M (1990) Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America **87**: 6286-6290.

Jornvall H (1975) Acetylation of Protein N-terminal amino groups structural observations on alpha-amino acetylated proteins. Journal of theoretical biology **55**: 1-12.

Jornvall H (1977) The primary structure of yeast alcohol dehydrogenase. European journal of biochemistry / FEBS **72**: 425-442.

Jornvall H, Fairwell T, Kratofil P, Wills C (1980) Differences in alpha-amino acetylation of isozymes of yeast alcohol dehydrogenase. FEBS letters **111**: 214-218.

Kamita M, Kimura Y, Ino Y, Kamp RM, Polevoda B, Sherman F, Hirano H (2011) N(alpha)-Acetylation of yeast ribosomal proteins and its effect on protein synthesis. Journal of proteomics **74**: 431-441.

Kayne PS, Kim UJ, Han M, Mullen JR, Yoshizaki F, Grunstein M (1988) Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. Cell **55**: 27-39.

Kimura M, Arndt E, Hatakeyama T, Kimura J (1989) Ribosomal proteins in halobacteria. Canadian journal of microbiology **35**: 195-199.

Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, Baginsky S (2007) Proteome dynamics during plastid differentiation in rice. Plant physiology **143**: 912-923.

Klussmann S, Franke P, Bergmann U, Kostka S, Wittmann-Liebold B (1993) N-terminal modification and amino-acid sequence of the ribosomal protein HmaS7 from *Haloarcula marismortui* and homology studies to other ribosomal proteins. Biological chemistry Hoppe-Seyler **374**: 305-312.

Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. The Journal of biological chemistry **284**: 478-485.

Kuo ML, den Besten W, Bertwistle D, Roussel MF, Sherr CJ (2004) N-terminal polyubiquitination and degradation of the Arf tumor suppressor. Genes & development **18**: 1862-1874.

Lee FJ, Lin LW, Smith JA (1989) Molecular cloning and sequencing of a cDNA encoding N alpha-acetyltransferase from *Saccharomyces cerevisiae*. The Journal of biological chemistry **264**: 12339-12343.

Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012 Jan; **40** (Database issue):D302-5. doi: 10.1093/nar/gkr931.

Liu LP, Deber CM (1998) Uncoupling hydrophobicity and helicity in transmembrane segments. Alpha-helical propensities of the amino acids in non-polar environments. The Journal of biological chemistry **273**: 23645-23648.

Liu LP, Deber CM (1999) Combining hydrophobicity and helicity: a novel approach to membrane protein structure prediction. Bioorganic & medicinal chemistry **7**: 1-7.

Mackay DT, Botting CH, Taylor GL, White MF (2007) An acetylase with relaxed specificity catalyses protein N-terminal acetylation in *Sulfolobus solfataricus*. Molecular microbiology **64**: 1540-1548.

Marmagne A, Ferro M, Meinnel T, Bruley C, Kuhn L, Garin J, Barbier-Brygoo H, Ephritikhine G (2007) A high content in lipid-modified peripheral proteins and integral receptor kinases features in the arabidopsis plasma membrane proteome. Molecular & cellular proteomics : MCP **6**: 1980-1996.

Marty F (1999) Plant vacuoles. The Plant cell **11**: 587-600.

Meyer AJ, Fricker MD (2000) Direct measurement of glutathione in epidermal cells of intact Arabidopsis roots by two-photon laser scanning microscopy. J Microsc **198**: 174-181.

Miao L, Fang H, Li Y, Chen H (2007) Studies of the in vitro N-alpha-acetyltransferase activities of E. coli RimL protein. Biochemical and biophysical research communications **357**: 641-647.

Midorikawa Y, Tsutsumi S, Taniguchi H, Ishii M, Kobune Y, Kodama T, Makuuchi M, Aburatani H (2002) Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis. Jpn J Cancer Res **93**: 636-643.

Mullen JR, Kayne PS, Moerschell RP, Tsunasawa S, Gribskov M, Colavito-Shepanski M, Grunstein M, Sherman F, Sternglanz R (1989) Identification and

characterization of genes and mutants for an N-terminal acetyltransferase from yeast. The EMBO journal **8**: 2067-2075.

Narita K (1958) Isolation of acetylpeptide from enzymic digests of TMV-protein. Biochimica et biophysica acta **28**: 184-191.

Neuwald AF, Landsman D (1997) GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. Trends in biochemical sciences **22**: 154-155.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng **10**: 1-6.

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci **12**: 357-358.

Pang AL, Clark J, Chan WY, Rennert OM (2011) Expression of human NAA11 (ARD1B) gene is tissue-specific and is regulated by DNA methylation. Epigenetics : official journal of the DNA Methylation Society **6**: 1391-1399.

Park EC, Szostak JW (1990) Point mutations in the yeast histone H4 gene prevent silencing of the silent mating type locus HML. Molecular and cellular biology **10**: 4932-4934.

Park EC, Szostak JW (1992) ARD1 and NAT1 proteins form a complex that has N-terminal acetyltransferase activity. The EMBO journal **11**: 2087-2093.

Perrot M, Massoni A, Boucherie H (2008) Sequence requirements for N-alpha-terminal acetylation of yeast proteins by NatA. Yeast **25**: 513-527.

Persson B, Flinta C, von Heijne G, Jornvall H (1985) Structures of N-terminally acetylated proteins. European journal of biochemistry / FEBS **152**: 523-527.

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

Polevoda B, Arnesen T, Sherman F (2009a) A synopsis of eukaryotic N-alpha-terminal acetyltransferases: nomenclature, subunits and substrates. BMC proceedings **4**: 1753-6561.

Polevoda B, Brown S, Cardillo TS, Rigby S, Sherman F (2008) Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. Journal of cellular biochemistry **103**: 492-508.

Polevoda B, Cardillo TS, Doyle TC, Bedi GS, Sherman F (2003) Nat3p and Mdm20p are required for function of yeast NatB N-alpha-terminal acetyltransferase and of actin and tropomyosin. The Journal of biological chemistry **278**: 30686-30697.

Polevoda B, Hoskins J, Sherman F (2009b) Properties of Nat4, an N(alpha)-acetyltransferase of Saccharomyces cerevisiae that modifies N termini of histones H2A and H4. Molecular and cellular biology **29**: 2913-2924.

Polevoda B, Norbeck J, Takakura H, Blomberg A, Sherman F (1999) Identification and specificities of N-terminal acetyltransferases from *Saccharomyces cerevisiae*. The EMBO journal **18**: 6155-6168.

Polevoda B, Sherman F (2001) NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. The Journal of biological chemistry **276**: 20154-20159.

Polevoda B, Sherman F (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. Journal of molecular biology **325**: 595-622.

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology **17**: 1030-1032.

Rope AF, Wang K, Evjenth R, Xing J, Johnston JJ, Swensen JJ, Johnson WE, Moore B, Huff CD, Bird LM, Carey JC, Opitz JM, Stevens CA, Jiang T, Schank C, Fain HD, Robison R, Dalley B, Chin S, South ST, Pysher TJ, Jorde LB, Hakonarson H, Lillehaug JR, Biesecker LG, Yandell M, Arnesen T, Lyon GJ (2011) Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. American journal of human genetics **89**: 28-43.

Rost B, Yachdav G, Liu J (2004) The PredictProtein server. Nucleic acids research **32**: W321-326.

Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K, Yamaguchi-Shinozaki K (2006) Dual function of an Arabidopsis transcription factor *DREB2A* in water-stress-responsive and heat-stress-responsive gene expression. Proceedings of the National Academy of Sciences of the United States of America **103**: 18822-18827.

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A. 1998 May **26;95(11)**:5857-64.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Oct **11;7**:539. doi: 10.1038/msb.2011.75.

Singer JM, Hermann GJ, Shaw JM (2000) Suppressors of mdm20 in yeast identify new alleles of ACT1 and TPM1 predicted to enhance actin-tropomyosin interactions. Genetics **156**: 523-534.

Singer JM, Shaw JM (2003) Mdm20 protein functions with Nat3 protein to acetylate Tpm1 protein and regulate tropomyosin-actin interactions in budding yeast. Proceedings of the National Academy of Sciences of the United States of America **100**: 7644-7649.

Smith ER, Eisen A, Gu W, Sattah M, Pannuti A, Zhou J, Cook RG, Lucchesi JC, Allis CD (1998) ESA1 is a histone acetyltransferase that is essential for growth in yeast. Proceedings of the National Academy of Sciences of the United States of America **95**: 3561-3565.

Song OK, Wang X, Waterborg JH, Sternglanz R (2003) An N-alpha-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A. The Journal of biological chemistry **278**: 38109-38112.

Soppa J (2010) Protein Acetylation in Archaea, Bacteria, and Eukaryotes. Archaea. **pii: 82068**. doi: 10.1155/2010/820681.

Starheim KK, Arnesen T, Gromyko D, Ryningen A, Varhaug JE, Lillehaug JR (2008) Identification of the human N(alpha)-acetyltransferase complex B (hNatB): a complex important for cell-cycle progression. The Biochemical journal **415**: 325-331.

Starheim KK, Gromyko D, Evjenth R, Ryningen A, Varhaug JE, Lillehaug JR, Arnesen T (2009a) Knockdown of human N-alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. Molecular and cellular biology **29**: 3569-3581.

Starheim KK, Gromyko D, Velde R, Varhaug JE, Arnesen T (2009b) Composition and biological significance of the human N-alpha-terminal acetyltransferases. BMC proceedings **3 Suppl 6**: S3.

Stephan I, (2011) Identifizierung und Charakterisierung von N-terminalen Acetyltransferasen in *Arabidopsis thaliana*. Dissertation.

Strous GJ, Berns AJ, Bloemendal H (1974) N-terminal acetylation of the nascent chains of alpha-crystallin. Biochemical and biophysical research communications **58**: 876-884.

Strous GJ, van Westreenen H, Bloemendal H (1973) Synthesis of lens protein in vitro. N-terminal acetylation of alpha-crystallin. European journal of biochemistry / FEBS **38**: 79-85.

Tanaka S, Matsushita Y, Yoshikawa A, Isono K (1989) Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12. Molecular & general genetics : MGG **217**: 289-293.

Tercero JC, Wickner RB (1992) MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly. The Journal of biological chemistry **267**: 20277-20281.

Timmermans MC, Maliga P, Vieira J, Messing J (1990) The pFF plasmids: cassettes utilising CaMV sequences for expression of foreign genes in plants.Journal of biotechnology **14**: 333-44.

Van Damme P, Arnesen T, Gevaert K (2011a) Protein alpha-N-acetylation studied by N-terminomics. The FEBS journal **278**: 3822-3834.

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

Van Damme P, Hole K, Pimenta-Marques A, Helsens K, Vandekerckhove J, Martinho RG, Gevaert K, Arnesen T (2011c) NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. PLoS genetics **7**: e1002169.

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

Van Damme P, Van Damme J, Demol H, Staes A, Vandekerckhove J, Gevaert K (2009) A review of COFRADIC techniques targeting protein N-terminal acetylation. BMC proceedings **3 Suppl 6**: S6.

Varshavsky A (2008) Discovery of cellular regulation by protein degradation. The Journal of biological chemistry **283**: 34469-34489.

Vetting MW, Bareich DC, Yu M, Blanchard JS (2008) Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. Protein science: a publication of the Protein Society **17**: 1781-1790.

Vetting MW, de Carvalho LP, Roderick SL, Blanchard JS (2005) A novel dimeric structure of the RimL N-alpha-acetyltransferase from Salmonella typhimurium. The Journal of biological chemistry **280**: 22108-22114.

Waterhouse, AM, Procter, JB, Martin, DMA, Clamp, M and Barton, GJ (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench". Bioinformatics **25 (9)** 1189-1191 doi: 10.1093/bioinformatics/btp033.

Whiteway M, Freedman R, Van Arsdell S, Szostak JW, Thorner J (1987) The yeast ARD1 gene product is required for repression of cryptic mating-type information at the HML locus. Molecular and cellular biology **7**: 3713-3722.

Whiteway M, Szostak JW (1985) The ARD1 gene of yeast functions in the switch between the mitotic cell cycle and alternative developmental pathways. Cell **43**: 483-492.

Williams BC, Garrett-Engele CM, Li Z, Williams EV, Rosenman ED, Goldberg ML (2003) Two putative acetyltransferases, san and deco, are required for establishing sister chromatid cohesion in Drosophila. Current biology : CB **13**: 2025-2036.

Wu FH, Shen SC, Lee LY, Lee SH, Chan MT, Lin CS (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. Plant Methods **5**: 1746-4811.

Yi CH, Pan H, Seebacher J, Jang IH, Hyberts SG, Heffron GJ, Vander Heiden MG, Yang R, Li F, Locasale JW, Sharfi H, Zhai B, Rodriguez-Mias R, Luithardt H, Cantley LC, Daley GQ, Asara JM, Gygi SP, Wagner G, Liu CF, Yuan J (2011) Metabolic regulation of protein N-alpha-acetylation by Bcl-xL promotes cell survival. Cell **146**: 607-620.

Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nature protocols **2**: 1565-1572.

Yoshikawa A, Isono S, Sheback A, Isono K (1987) Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12. Molecular & general genetics : MGG **209**: 481-488.

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

Table S1A. List of acetylated proteins in pNaa30p-expressed *E. coli* characterized by LC-MS/MS.

No.	Accession	aa before	N-termini	%Ac control	%Ac sample	Description
1	P0A8L7	-	MDMDLNNRLT	0%	38%	UPF0263 protein yciU
2	P00579	-	MEQNPQSQLK	1%	30%	RNA polymerase sigma factor rpoD
3	P0A8M6	-	METTKPSFQD	4%	11%	UPF0265 protein yeeX
4	P0AFD1	-	MHENQQPQTE	1%	33%	NADH-quinone oxidoreductase subunit E
5	P0ADY3	-	MIQEQTMLNV	1%	9%	50S ribosomal protein L14
6	P0AEX3	M	AESTVTADS	0%	55%	Alpha-ketoglutarate permease
7	P0A729	M	PKLILASTS	1%	27%	Maf-like protein yceF
8	P21170	M	SDDMSMGLP	0%	99%	Biosynthetic arginine decarboxylase
9	P0A8N3	M	SEQHAQGAD	34%	63%	Lysine--tRNA ligase
10	P32144	M	SLTELTGNP	6%	14%	Uncharacterized HTH-type transcriptional regulator YihW
11	P0ACK2	M	SNTDASGEK	5%	10%	Putative aga operon transcriptional repressor
12	P09832	M	SQNVYQFID	0%	36%	Glutamate synthase [NADPH] small chain
13	P40191	M	SSLLLFNDK	75%	94%	Pyridoxine kinase
14	P33599	M	TDLTAQEPA	0%	64%	NADH-quinone oxidoreductase subunit C/D
15	P0ACM2	M	TVETQLNPT	23%	54%	Uncharacterized HTH-type transcriptional regulator YdfH

Table S1B. List of acetylated proteins in pNaa10p-expressed *E. coli* characterized by LC-MS/MS.

No.	Accession	aa before	N-termini	%Ac control	%Ac sample	Description
1	P0ACY1	-	MDALELLINR	1%	69%	Putative NAD(P)H nitroreductase ydjA
2	F2X7N5	-	MDIISVALKR	1%	32%	Nitroreductase
3	P64559	-	MDINNKARIH	0%	64%	UPF0350 protein ygfY
4	P0AFZ3	-	MDLSQLTPRR	27%	98%	Stringent starvation protein B
5	P0A8L7	-	MDMDLNNRLT	0%	100%	UPF0263 protein yciU
6	P0ACT6	-	MMDNMQTEAQ	0%	100%	HTH-type transcriptional regulator UidR
7	P0ACI0	-	MDQAGIIRDL	1%	61%	Right origin-binding protein
8	P0AFC7	-	MDYTLTRIDP	1%	60%	NADH-quinone oxidoreductase subunit B
9	P15038	-	MELKATTLGK	0%	97%	Helicase IV
10	P0A853	-	MENFKHLPEP	0%	30%	Tryptophanase
11	P00579	-	MEQNPQSQLK	1%	100%	RNA polymerase sigma factor rpoD
12	P0AEW4	-	MESLLTLPLA	4%	60%	3',5'-cyclic adenosine monophosphate phosphodiesterase CpdA
13	P0AAD6	-	METTQTSTIA	0%	99%	Serine transporter
14	P0AGD7	-	MFDNLTDRLS	0%	94%	Signal recognition particle protein
15	P07650	-	MFLAQEIIRK	1%	19%	Thymidine phosphorylase
16	P0AFD1	-	MHENQQPQTE	1%	96%	NADH-quinone oxidoreductase subunit E
17	P64540	-	MIAEFESRIL	0%	95%	Uncharacterized protein yfcL
18	P0AB20	-	MIASKFGIGQ	5%	83%	Heat shock protein HspQ
19	P22333	-	MIDTTLPLTD	0%	32%	Adenosine deaminase
20	P33232	-	MIISAASDYR	2%	37%	L-lactate dehydrogenase [cytochrome]

21	P0ABH0	-	MIKATDRKLV	0%	24%	Cell division protein FtsA
22	P0ADY3	-	MIQEQTMLNV	1%	22%	50S ribosomal protein L14
23	P0A7C2	-	MKALTARQQE	0%	18%	LexA repressor
24	P0A9M2	-	MKHTVEVMIP	4%	31%	Hypoxanthine phosphoribosyltransferase
25	P0ACP1	-	MKLDEIARLA	0%	45%	Catabolite repressor/activator
26	P0A9S3	-	MKSVVNDTDG	1%	28%	Galactitol-1-phosphate 5-dehydrogenase
27	P0A8L1	-	MLDPNLLRNE	0%	35%	Serine--tRNA ligase
28	P0A9X4	-	MLKKFRGMFS	0%	27%	Rod shape-determining protein MreB
29	P06989	-	MLTEQQRREL	0%	54%	Histidine biosynthesis bifunctional protein HisIE
30	P77737	-	MNAVTEGRKV	0%	96%	Oligopeptide transport ATP-binding protein OppF
31	P0AF08	-	MNEQSQAKSP	2%	99%	Protein mrp OS=Escherichia coli (strain K12) GN=mrp PE=3 SV=1
32	P00864	-	MNEQYSALRS	0%	93%	Phosphoenolpyruvate carboxylase
33	P61714	-	MNIIEANVAT	0%	45%	6,7-dimethyl-8-ribityllumazine synthase
34	P0AB61	-	MNKETQPIDR	1%	100%	Protein yciN
35	P27434	-	MNTEATHDQN	0%	99%	Cytoskeleton protein rodZ
36	P17952	-	MNTQQLAKLR	0%	95%	UDP-N-acetylmuramate--L-alanine ligase
37	P0AB91	-	MNYQNDDLRI	0%	32%	Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive
38	P0A7Z4	-	MQGSVTEFLK	4%	32%	DNA-directed RNA polymerase subunit alpha
39	P0A7D7	-	MQKQAELYRG	0%	61%	Phosphoribosylaminoimidazole-succinocarboxamide synthase
40	P0ABB0	-	MQLNSTEISE	1%	50%	ATP synthase subunit alpha
41	P0A850	-	MQVSVETTQG	3%	18%	Trigger factor
42	P0A6A8	-	MSTIEERVKK	0%	30%	Acyl carrier protein
43	P0A9J8-a	-	MTSENPLLAL	0%	86%	P-protein
44	P0AC44-a	-	MVSNASALGR	0%	93%	Succinate dehydrogenase hydrophobic membrane anchor subunit
45	P63224	-	MYQDLIRNEL	1%	28%	Phosphoheptose isomerase
46	P0ACL2	-	MEITEPRRLY	0%	100%	Exu regulon transcriptional regulator

47	P60723	-	MELVLKDAQS	2%	12%	50S ribosomal protein L4
48	P0A8W8	-	MEMTNAQRLI	1%	13%	UPF0304 protein yfbU
49	P21151	-	MEQVVIVDAI	0%	57%	3-ketoacyl-CoA thiolase
50	P07012	-	MFEINPVNNR	1%	13%	Peptide chain release factor 2
51	P0ADK8	-	MKEVEKNEIK	4%	11%	Uncharacterized protein yibL
52	P0A6Z3	-	MKGQETRGFQ	1%	18%	Chaperone protein htpG
53	P0AC41	-	MKLPVREFDA	0%	8%	Succinate dehydrogenase flavoprotein subunit
54	P00963	-	MKTAYIAKQR	0%	9%	Aspartate--ammonia ligase
55	P0C8J8	-	MKTLIARHKA	0%	19%	D-tagatose-1,6-bisphosphate aldolase subunit gatZ
56	P75821	-	MKVLVTGATS	2%	8%	Uncharacterized protein ybjS
57	P69913	-	MLILTRRVGE	0%	13%	Carbon storage regulator
58	P13016	-	MLLEQGWLVG	0%	65%	1,6-anhydro-N-acetylmuramyl-L-alanine amidase AmpD
59	P0AFX4	-	MLNQLDNLTE	0%	89%	Regulator of sigma D
60	P75990	-	MLTTLIYRSH	0%	87%	Blue light- and temperature-regulated antirepressor YcgF
61	P76469	-	MNALLSNPFK	0%	72%	2-keto-3-deoxy-L-rhamnonate aldolase
62	P0A9K3	-	MNIDTREITL	0%	87%	PhoH-like protein
63	P0AAT9	-	MNKVAQYYRE	0%	28%	Uncharacterized protein ybeL
64	P0A836	-	MNLHEYQAKQ	4%	17%	Succinyl-CoA ligase [ADP-forming] subunit beta
65	P62623	-	MQILLANPRG	1%	17%	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
66	P77569	-	MQNNEQTEYK	0%	99%	Mhp operon transcriptional activator
67	P0A9D8	-	MQQLQNIIET	1%	10%	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
68	P77712	-	MQTQIKVRGY	0%	100%	Long-chain acyl-CoA thioesterase FadM
69	P52074	-	MQTQLTEEMR	0%	100%	Glycolate oxidase iron-sulfur subunit
70	P60716	-	MSKPIVMERG	0%	6%	Lipoyl synthase
71	P25524	-	MSNNALQTII	0%	9%	Cytosine deaminase
72	P16703	-	MSTLEQTIGN	2%	10%	Cysteine synthase B

73	P0A6Y1	-	MTKSELIERL	1%	12%	Integration host factor subunit beta
74	P0A6E4	-	MTTILKHLPV	0%	8%	Argininosuccinate synthase
75	P0A9H1	-	MVEDILAPGL	3%	10%	G/U mismatch-specific DNA glycosylase
76	P0A8V2	-	MVYSYTEKKR	1%	13%	DNA-directed RNA polymerase subunit beta
77	P0AEP3	M	AAINTKVKK	2%	44%	UTP--glucose-1-phosphate uridylyltransferase
78	P0ACE7	M	AEETIFSKI	0%	66%	HIT-like protein hinT
79	P0A6P1	M	AEITASLVK	1%	38%	Elongation factor Ts
80	P0C093	M	AEKQTAKRN	1%	91%	Nucleoid occlusion factor SlmA
81	P0ADN2	M	AESFTTTNR	0%	100%	UPF0438 protein yifE
82	P0AEX3	M	AESTVTADS	0%	97%	Alpha-ketoglutarate permease
83	P0ADE8	M	AFTPFPPRQ	0%	74%	tRNA-modifying protein ygfZ
84	P0ACB0	M	AGNKPFNKQ	4%	92%	Replicative DNA helicase
85	P0A7G6	M	AIDENKQKA	1%	87%	Protein RecA
86	P18335	M	AIEQTAITR	0%	53%	Acetylornithine/succinyldiaminopimelate aminotransferase
87	P10121	M	AKEKKRGFF	0%	100%	Signal recognition particle receptor FtsY
88	P0AAU7	M	MAKEQTDRTT	0%	45%	Uncharacterized protein ybfE
89	P51024	M	AKLTLQEQL	3%	58%	Uncharacterized protein yaiL
90	P63228	M	AKSVPAIFL	1%	31%	D,D-heptose 1,7-bisphosphate phosphatase
91	P0A734	M	ALLDFFLSR	1%	87%	Cell division topological specificity factor
92	P0A6Z1	M	ALLQISEPG	0%	98%	Chaperone protein HscA
93	P0ABN1	M	ANNTTGFTR	0%	99%	Diacylglycerol kinase
94	P0AF90	M	ANPEQLEEQ	1%	90%	Regulator of ribonuclease activity B
95	P03024	M	ATIKDVARL	1%	45%	HTH-type transcriptional regulator GalR
96	P0A9V1	M	ATLTAKNLA	0%	25%	Lipopolysaccharide export system ATP-binding protein LptB
97	P0A7E9	M	ATNAKPVYK	1%	33%	Uridylate kinase
98	P0ACA3	M	AVAANKRSV	0%	100%	Stringent starvation protein A

99	P62707	M	AVTKLVLVR	0%	13%	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase
100	P0AC53	M	AVTQTAQAC	3%	36%	Glucose-6-phosphate 1-dehydrogenase
101	P0AAJ5	M	AYQSQDIIR	1%	97%	Formate dehydrogenase-O iron-sulfur subunit
102	P0ACD4	M	AYSEKVIDH	0%	99%	NifU-like protein
103	P0ACL9	M	AYSKIRQPK	0%	78%	Pyruvate dehydrogenase complex repressor
104	P09546	M	GTTTMGVKL	6%	19%	Bifunctional protein putA
105	P21170	M	SDDMSMGLP	0%	96%	Biosynthetic arginine decarboxylase
106	P0AFG0	M	SEAPKKRWY	1%	59%	Transcription antitermination protein NusG
107	P63177	M	SEMIYGIHA	6%	16%	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB
108	P0A8N3	M	SEQHAQGAD	34%	88%	Lysine--tRNA ligase
109	P77536	M	SIKTSNTDF	1%	32%	Uncharacterized electron transport protein ykgF
110	P0ABK5	M	SKIFEDNSL	3%	29%	Cysteine synthase A
111	P0A744	M	SLFDKKHLV	2%	52%	Peptide methionine sulfoxide reductase MsrA
112	P0ADZ4	M	SLSTEATAK	1%	16%	30S ribosomal protein S15
113	P32144	M	SLTELTGNP	6%	98%	Uncharacterized HTH-type transcriptional regulator YihW
114	P31801	M	SNAQEAVKT	1%	71%	Calcium/proton antiporter
115	P0AF28	M	SNQEPATIL	0%	81%	Nitrate/nitrite response regulator protein NarL
116	P0AES6	M	SNSYDSSSI	5%	94%	DNA gyrase subunit B
117	P0ACK2	M	SNTDASGEK	5%	100%	Putative aga operon transcriptional repressor
118	P21499	M	SQDPFQERE	0%	92%	Ribonuclease R
119	P09832	M	SQNVYQFID	0%	95%	Glutamate synthase [NADPH] small chain
120	P0AEZ9	M	SQVSTEFIP	3%	98%	Molybdenum cofactor biosynthesis protein B
121	P0AC69	M	STTIEKIQR	0%	19%	Glutaredoxin-4
122	P00562	M	SVIAQAGAK	5%	100%	Bifunctional aspartokinase/homoserine dehydrogenase 2
123	P05020	M	TAPSQVLKI	0%	50%	Dihydroorotase
124	P0A991	M	TDIAQLLGK	3%	21%	Fructose-bisphosphate aldolase class 1

125	P33599	M	TDLTAQEPA	0%	67%	NADH-quinone oxidoreductase subunit C/D
126	P37095	M	TEAMKITLS	1%	44%	Peptidase B
127	P0A8R7	M	TEPLKPRID	0%	29%	UPF0283 membrane protein YcjF
128	P09053	M	TFSLFGDKF	8%	30%	Valine--pyruvate aminotransferase
129	P0A7I4	M	TLSPYLQEV	2%	21%	Peptide chain release factor 3
130	P08244	M	TLTASSSSR	2%	100%	Orotidine 5'-phosphate decarboxylase
131	B8LFD6	M	TMITDSLAV	3%	62%	Beta-galactosidase 2
132	P0AD49	M	TMNITSKQM	1%	49%	Ribosome-associated inhibitor A
133	P69828	M	TNLFVRSGI	0%	95%	Galactitol-specific phosphotransferase enzyme IIA component
134	P69829	M	TNNDTTLQL	3%	94%	Nitrogen regulatory protein
135	P04825	M	TQQPQAKYR	2%	36%	Aminopeptidase N
136	P75838	M	TQTFIPGKD	2%	98%	UPF0142 protein ycaO
137	P33195	M	TQTLSQLEN	1%	56%	Glycine dehydrogenase [decarboxylating]
138	P0A9J8-b	M	TSENPLLAL	4%	68%	P-protein
139	P39160	M	TTIVDSNLP	1%	81%	D-mannonate oxidoreductase
140	P0A7B8	M	TTIVSVRRN	0%	20%	ATP-dependent protease subunit HslV
141	P0ACM2	M	TVETQLNPT	23%	95%	Uncharacterized HTH-type transcriptional regulator YdfH
142	P05706	M	TVIYQTTIT	0%	100%	Glucitol/sorbitol-specific phosphotransferase enzyme IIA component
143	P0A7Y8	M	VKLAFPREL	4%	35%	Ribonuclease P protein component
144	P0AC44-b	M	VSNASALGR	0%	75%	Succinate dehydrogenase hydrophobic membrane anchor subunit
145	P06721	M	ADKKLDTQL	5%	11%	Cystathionine beta-lyase metC
146	P0A794	M	AELLLGVNI	67%	20%	Pyridoxine 5'-phosphate synthase
147	P12282	M	AELSDQEML	0%	92%	Molybdopterin-synthase adenylyltransferase
148	P0AFN2	M	AGINLNKKL	0%	100%	Phage shock protein C
149	P12008	M	AGNTIGQLF	0%	17%	Chorismate synthase
150	P0A6X3	M	AKGQSLQDP	1%	12%	Protein hfq

151	P0A7J3	M	ALNLQDKQA	0%	13%	50S ribosomal protein L10
152	P27248	M	AQQTPLYEQ	1%	16%	Aminomethyltransferase
153	P64479	M	ASGDLVRYV	0%	86%	Uncharacterized protein ydiZ
154	P0AGJ9	M	ASSNLIKQL	1%	8%	Tyrosine--tRNA ligase
155	P75952	M	ATDSTQCVK	0%	94%	Uncharacterized HTH-type transcriptional regulator YcfQ
156	P29013	M	ATIDSMNKD	0%	100%	Uncharacterized protein ycgB
157	P0ACP7	M	ATIKDVAKR	0%	12%	HTH-type transcriptional repressor PurR
158	P22634	M	ATKLQDGNT	0%	55%	Glutamate racemase
159	P0A6N4	M	ATYYSNDFR	1%	17%	Elongation factor P
160	P60422	M	AVVKCKPTS	1%	7%	50S ribosomal protein L2
161	P75713	M	GYLNNVTGY	0%	50%	Uncharacterized protein ylbA
162	P0AED7	M	SCPVIELTQ	0%	12%	Succinyl-diaminopimelate desuccinylase
163	P0A9U6	M	SDEGLAPGK	0%	84%	HTH-type transcriptional regulator PuuR
164	P04951	M	SFVVIIPAR	0%	6%	3-deoxy-manno-octulosonate cytidylyltransferase
165	P16676	M	SIEIANIKK	0%	22%	Sulfate/thiosulfate import ATP-binding protein CysA
166	P23893	M	SKSENLYSA	1%	18%	Glutamate-1-semialdehyde 2,1-aminomutase
167	P37330	M	SQTITQSRL	0%	68%	Malate synthase G
168	P00550	M	SSDIKIKVQ	0%	93%	PTS system mannitol-specific EIICBA component
169	P40191	M	SSLLLFNDK	75%	100%	Pyridoxine kinase
170	P0ABT2	M	STAKLVKSK	7%	17%	DNA protection during starvation protein
171	P60651	M	STLGHQYDN	0%	42%	Agmatinase
172	P77549	M	TAVSQTETR	0%	100%	UPF0226 protein yfcJ
173	P0A9A9	M	TDNNTALKK	2%	18%	Ferric uptake regulation protein
174	P0A705	M	TDVTIKTLA	3%	14%	Translation initiation factor IF-2
175	P0A7F3	M	THDNKLQVE	0%	72%	Aspartate carbamoyltransferase regulatory chain
176	P23843	M	TNITKRSLV	0%	95%	Periplasmic oligopeptide-binding protein

177	P04335	M	TQANLSETL	0%	99%	Esterase FrsA
178	P0ACU2	M	TQGAVKTTG	0%	76%	HTH-type transcriptional regulator RutR
179	P29012	M	TRPIQASLD	0%	16%	Alanine racemase, catabolic
180	P0AEN1	M	TTLSCKVTS	0%	14%	NAD(P)H-flavin reductase
181	P38038	M	TTQVPPSAL	0%	99%	Sulfite reductase [NADPH] flavoprotein alpha-component
182	P16528	M	VAPIPAKRG	0%	99%	Acetate operon repressor
183	P0ACJ0	M	VDSKKRPGK	64%	93%	Leucine-responsive regulatory protein

Table S1C. List of acetylated proteins in pNaa20p-expressed *E. coli* characterized by LC-MS/MS.

No.	Accession	aa before	N-termini	%Ac control	%Ac sample	Description
1	F2X7N5	-	MDIISVALKR	1%	80%	Nitroreductase
2	P00579	-	MEQNPQSQLK	1%	70%	RNA polymerase sigma factor rpoD
3	P0A8M6	-	METTKPSFQD	4%	94%	UPF0265 protein yeeX
4	P0AAD6	-	METTQTSTIA	0%	61%	Serine transporter
5	P0AGD7	-	MFDNLTDRLS	0%	94%	Signal recognition particle protein
6	P07012	-	MFEINPVNNR	1%	73%	Peptide chain release factor 2
7	P0ABB8	-	MFKEIFTRLI	0%	100%	Magnesium-transporting ATPase, P-type 1 O
8	P07650	-	MFLAQEIIRK	1%	99%	Thymidine phosphorylase
9	P30850	-	MFQDNPLLAQ	0%	85%	Exoribonuclease 2
10	P68919	-	MFTINAEVRK	3%	55%	50S ribosomal protein L25
11	P64540	-	MIAEFESRIL	0%	72%	Uncharacterized protein yfcL
12	Q46868	-	MIDPKKIEQI	2%	76%	Uncharacterized protein yqiC
13	P22333	-	MIDTTLPLTD	0%	37%	Adenosine deaminase
14	P10371	-	MIIPALDLID	0%	61%	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase
15	P33232	-	MIISAASDYR	2%	54%	L-lactate dehydrogenase [cytochrome]
16	P0ADY3	-	MIQEQTMLNV	1%	32%	50S ribosomal protein L14
17	P0AA10	-	MKTFTAKPET	1%	35%	50S ribosomal protein L13
18	P0A8L1	-	MLDPNLLRNE	0%	44%	Serine--tRNA ligase
19	P31663	-	MLIIETLPLL	0%	57%	Pantothenate synthetase
20	P69913	-	MLILTRRVGE	0%	64%	Carbon storage regulator

21	P21888	-	MLKIFNTLTR	0%	36%	Cysteine--tRNA ligase
22	P0A9X4	-	MLKKFRGMFS	0%	92%	Rod shape-determining protein MreB
23	P64606	-	MLLNALASLG	0%	86%	Probable phospholipid ABC transporter permease protein mlaE
24	P05055	-	MLNPIVRKFQ	0%	40%	Polyribonucleotide nucleotidyltransferase
25	P0AAA1	-	MNIFEQTPPN	0%	99%	Inner membrane protein yagU
26	P61714	-	MNIIEANVAT	0%	51%	6,7-dimethyl-8-ribityllumazine synthase
27	P27434	-	MNTEATHDQN	0%	23%	Cytoskeleton protein rodZ
28	P17952	-	MNTQQLAKLR	0%	85%	UDP-N-acetylmuramate--L-alanine ligase
29	P62623	-	MQILLANPRG	1%	61%	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
30	P0AEK4	-	MGFLSGKRIL	0%	36%	Enoyl-[acyl-carrier-protein] reductase [NADH] FabI
31	P0A6A8	-	MSTIEERVKK	0%	87%	Acyl carrier protein
32	P0A8M0	-	MSVVPVADVL	0%	21%	Asparagine--tRNA ligase
33	P0AC44	-	MVSNASALGR	0%	60%	Succinate dehydrogenase hydrophobic membrane anchor subunit
34	P0ACY1	-	MDALELLINR	1%	17%	Putative NAD(P)H nitroreductase ydjA
35	P0A853	-	MENFKHLPEP	0%	11%	Tryptophanase
36	P0C8J8	-	MKTLIARHKA	0%	9%	D-tagatose-1,6-bisphosphate aldolase subunit gatZ
37	P24182	-	MLDKIVIANR	1%	14%	Biotin carboxylase
38	P13016	-	MLLEQGWLVG	0%	65%	1,6-anhydro-N-acetylmuramyl-L-alanine amidase AmpD
39	P75990	-	MLTTLIYRSH	0%	28%	Blue light- and temperature-regulated antirepressor YcgF
40	P0ABB0	-	MQLNSTEISE	1%	16%	ATP synthase subunit alpha
41	P0A7V0	-	MATVSMRDML	0%	100%	30S ribosomal protein S2
42	P0AG48	-	MYAVFQSGGK	0%	13%	50S ribosomal protein L21
43	P0ADN2	M	AESFTTTNR	0%	78%	UPF0438 protein yifE
44	P0AEX3	M	AESTVTADS	0%	57%	Alpha-ketoglutarate permease
45	P0ADE8	M	AFTPFPPRQ	0%	78%	tRNA-modifying protein ygfZ
46	P0A7G6	M	AIDENKQKA	1%	31%	Protein RecA

47	P18335	M	AIEQTAITR	0%	19%	Acetylornithine/succinyldiaminopimelate aminotransferase
48	P51024	M	AKLTLQEQL	3%	76%	Uncharacterized protein yaiL
49	P0AG51	M	AKTIKITQT	1%	36%	50S ribosomal protein L30
50	P0A734	M	ALLDFFLSR	1%	97%	Cell division topological specificity factor
51	P0A6Z1	M	ALLQISEPG	0%	96%	Chaperone protein HscA
52	P05793	M	ANYFNTLNL	0%	26%	Ketol-acid reductoisomerase
53	P0AG96	M	SANTEAQGS	0%	89%	Preprotein translocase subunit SecE
54	P21170	M	SDDMSMGLP	0%	93%	Biosynthetic arginine decarboxylase
55	P04951	M	SFVVIIPAR	0%	37%	3-deoxy-manno-octulosonate cytidylyltransferase
56	P0A742	M	SIIKEFREF	1%	22%	Large-conductance mechanosensitive channel
57	P0AAC8	M	SITLSDSAA	2%	63%	Iron-binding protein IscA
58	P33920	M	SLDINQIAL	0%	52%	Nucleoid-associated protein YejK
59	P0A744	M	SLFDKKHLV	2%	58%	Peptide methionine sulfoxide reductase MsrA
60	P0ADZ4	M	SLSTEATAK	1%	25%	30S ribosomal protein S15
61	P31801	M	SNAQEAVKT	1%	99%	Calcium/proton antiporter
62	P0AEH5	M	SNQFGDTRI	0%	88%	Protein elaB
63	P0AES6	M	SNSYDSSSI	5%	23%	DNA gyrase subunit B
64	P0ACG1	M	SVMLQSLNN	4%	50%	DNA-binding protein stpA
65	P0A890	M	TDLFSSPDH	0%	50%	Sulfurtransferase TusA
66	P33599	M	TDLTAQEPA	0%	39%	NADH-quinone oxidoreductase subunit C/D
67	P0AG67	M	TESFAQLFE	5%	94%	30S ribosomal protein S1
68	P09053	M	TFSLFGDKF	8%	100%	Valine--pyruvate aminotransferase
69	P08244	M	TLTASSSSR	2%	81%	Orotidine 5'-phosphate decarboxylase
70	P0AD49	M	TMNITSKQM	1%	100%	Ribosome-associated inhibitor A
71	P75838	M	TQTFIPGKD	2%	83%	UPF0142 protein ycaO
72	P0AC44	M	VSNASALGR	0%	94%	Succinate dehydrogenase hydrophobic membrane anchor subunit

73	P06721	M	ADKKLDTQL	5%	34%	Cystathionine beta-lyase metC
74	P0A794	M	AELLLGVNI	3%	31%	Pyridoxine 5'-phosphate synthase
75	P0AFN2	M	AGINLNKKL	0%	100%	Phage shock protein C
76	P0AE18	M	AISIKTPED	2%	25%	Methionine aminopeptidase
77	P69222	M	AKEDNIEMQ	3%	16%	Translation initiation factor IF-1
78	P0A7G2	M	AKEFGRPQR	0%	25%	Ribosome-binding factor A
79	P10121	M	AKEKKRGFF	0%	16%	Signal recognition particle receptor FtsY
80	P0A9V1	M	ATLTAKNLA	0%	31%	Lipopolysaccharide export system ATP-binding protein LptB
81	P0A6W0	M	AVAMDNAIL	0%	85%	Glutaminase 2
82	P17115	M	SEALLNAGR	1%	18%	Arabinose 5-phosphate isomerase GutQ
83	P63177	M	SEMIYGIHA	6%	13%	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB
84	P77488	M	SFDIAKYPT	3%	19%	1-deoxy-D-xylulose-5-phosphate synthase
85	P25888	M	SFDSLGLSP	0%	83%	ATP-dependent RNA helicase rhlE
86	P39286	M	SKNKLSKGQ	1%	10%	Putative ribosome biogenesis GTPase RsgA
87	P23893	M	SKSENLYSA	1%	10%	Glutamate-1-semialdehyde 2,1-aminomutase
88	P03004	M	SLSLWQQCL	0%	99%	Chromosomal replication initiator protein DnaA
89	P0A7W7	M	SMQDPIADM	1%	14%	30S ribosomal protein S8
90	P09832	M	SQNVYQFID	0%	48%	Glutamate synthase [NADPH] small chain
91	P37330	M	SQTITQSRL	0%	61%	Malate synthase G
92	P40191	M	SSLLLFNDK	75%	96%	Pyridoxine kinase
93	P0ABT2	M	STAKLVKSK	7%	14%	DNA protection during starvation protein
94	P06986	M	STVTITDLA	1%	15%	Histidinol-phosphate aminotransferase
95	P77549	M	TAVSQTETR	0%	97%	UPF0226 protein yfcJ
96	P0A9B2	M	TIKVGINGF	1%	16%	Glyceraldehyde-3-phosphate dehydrogenase A
97	P0A7I4	M	TLSPYLQEV	2%	27%	Peptide chain release factor 3
98	B8LFD6	M	TMITDSLAV	3%	14%	Beta-galactosidase 2

99	P0A7B8	M	TTIVSVRRN	0%	12%	ATP-dependent protease subunit HslV
100	P0ACM2	M	TVETQLNPT	23%	38%	Uncharacterized HTH-type transcriptional regulator YdfH
101	P0ABA6	M	AGAKEIRSK	0%	10%	ATP synthase gamma chain

Table S2. Constructs were used and created in this work

Vector	Bacterial resistance	Plant selection	Description
pFF19-EYFP-Naa40	Amp	-	Naa40-EYFP C-terminal fusion, *KpnI* and *XbaI* (Timmermans et al., 1990)
pFF19-EYFP-Naa60	Amp	-	Naa60-EYFP C-terminal fudion, *BamHI* and *SalI*
pFF19-EYFP-At1g24040	Amp	-	At1g24040-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At2g39000	Amp	-	At2g39000-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At2g06025	Amp	-	At2g06025-EYFP C-terminal fusion, *BamHI* and *XbaI*
pFF19-EYFP-At1g32070	Amp	-	At1g32070-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At1g26220	Amp	-	At1g26220-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At4g19985	Amp	-	At4g19985-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At4g28030	Amp	-	At4g28030-EYFP C-terminal fusion, *BamHI* and *XbaI*
pFF19-EYFP-At1g03650	Amp	-	At1g03650-EYFP C-terminal fusion, *BamHI* and *XbaI*
pFF19-EYFP-At3g02980	Amp	-	At3g02980-EYFP C-terminal fusion, *BamHI* and *SalI*
pFF19-EYFP-At4g72030	Amp	-	At4g72030-EYFP C-terminal fusion, *BamHI* and *XbaI*
pB7YWG2-Naa60 (p35S:AtNaa60-YFP)	Spec	Basta	Naa60-EYFP C-terminal fusion, Gateway
pB7WGY2-Naa60 (p35S:YFP-AtNaa60)	Spec	Basta	EYFP-Naa60 N-terminal fusion, Gateway
pETM41-Naa40	Kan	-	MBP-(His)6- Naa40 for expression and purification, *NcoI* and *KpnI*
pETM41-Naa60	Kan	-	MBP-(His)6- Naa60 for expression and purification, *NcoI* and

			BamHI
pETM41-At1g24040	Kan	-	MBP-(His)6- At1g24040 for expression and purification, *NcoI* and *BamHI*
pETM41-At2g39000	Kan	-	MBP-(His)6- At2g39000 for expression and purification, *NcoI* and *BamHI*
pETM41-At2g06025	Kan	-	MBP-(His)6- At2g06025 for expression and purification, *NcoI* and *BamHI*
pURT2-Ubi-RFP	Spec	Kan	RFP cytosolic marker, provided by AG Schumacher, COS Heidelberg
pKF13-TMD23-RFP	Amp	-	RFP plasma membrane marker, provided by AG Schumacher, COS

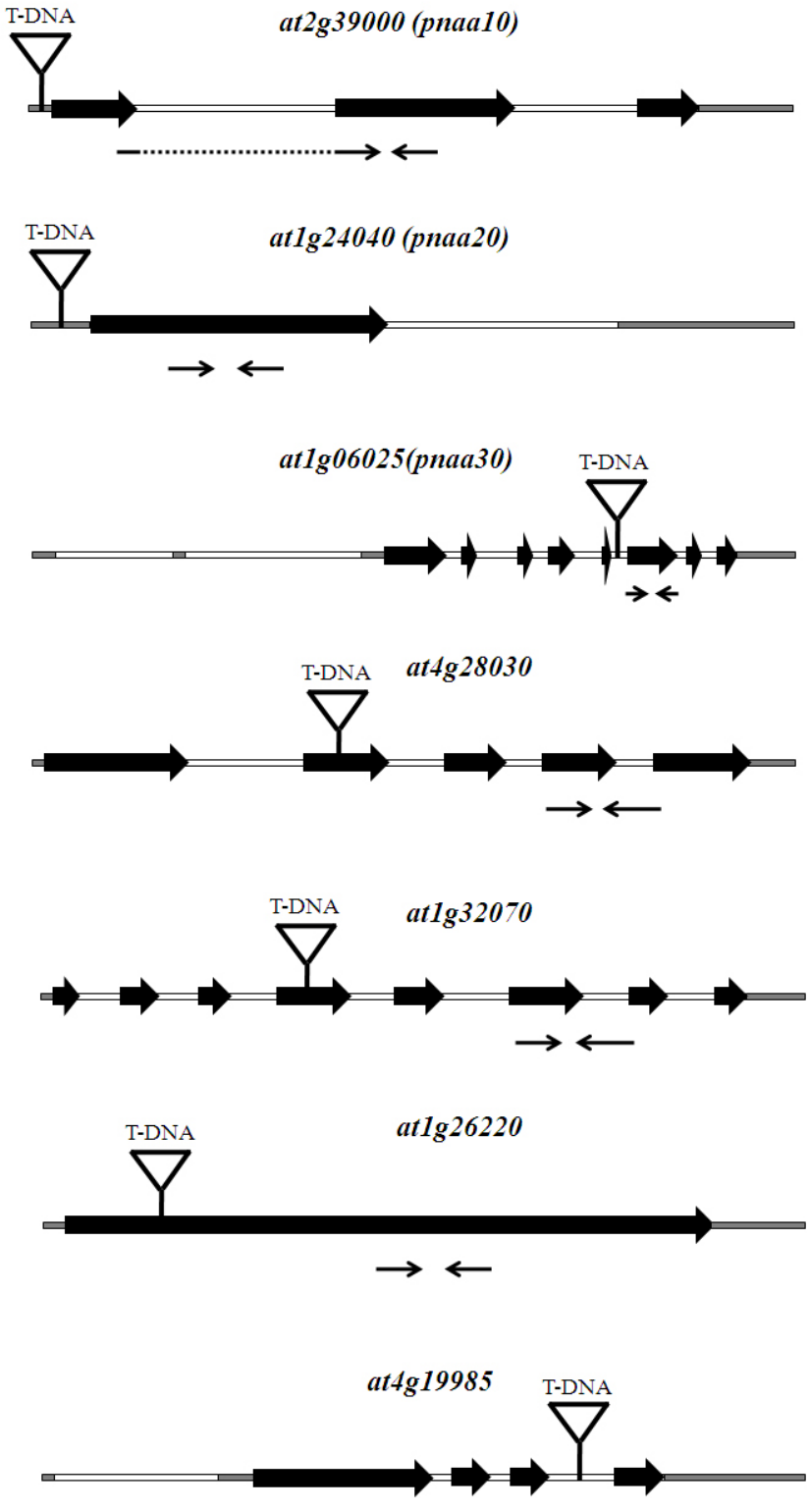


Figure S1. Strucsture of different pNAT genes with T-DNA insertions.

T-DNA insertion was inserted at the different positions at seven genes coding putative pNATs. Big black arrow indicates exon, white bar indicates intron, grey bar indicates un-translated region (UTR) and small back arrow indicates qRT-PCR primers.

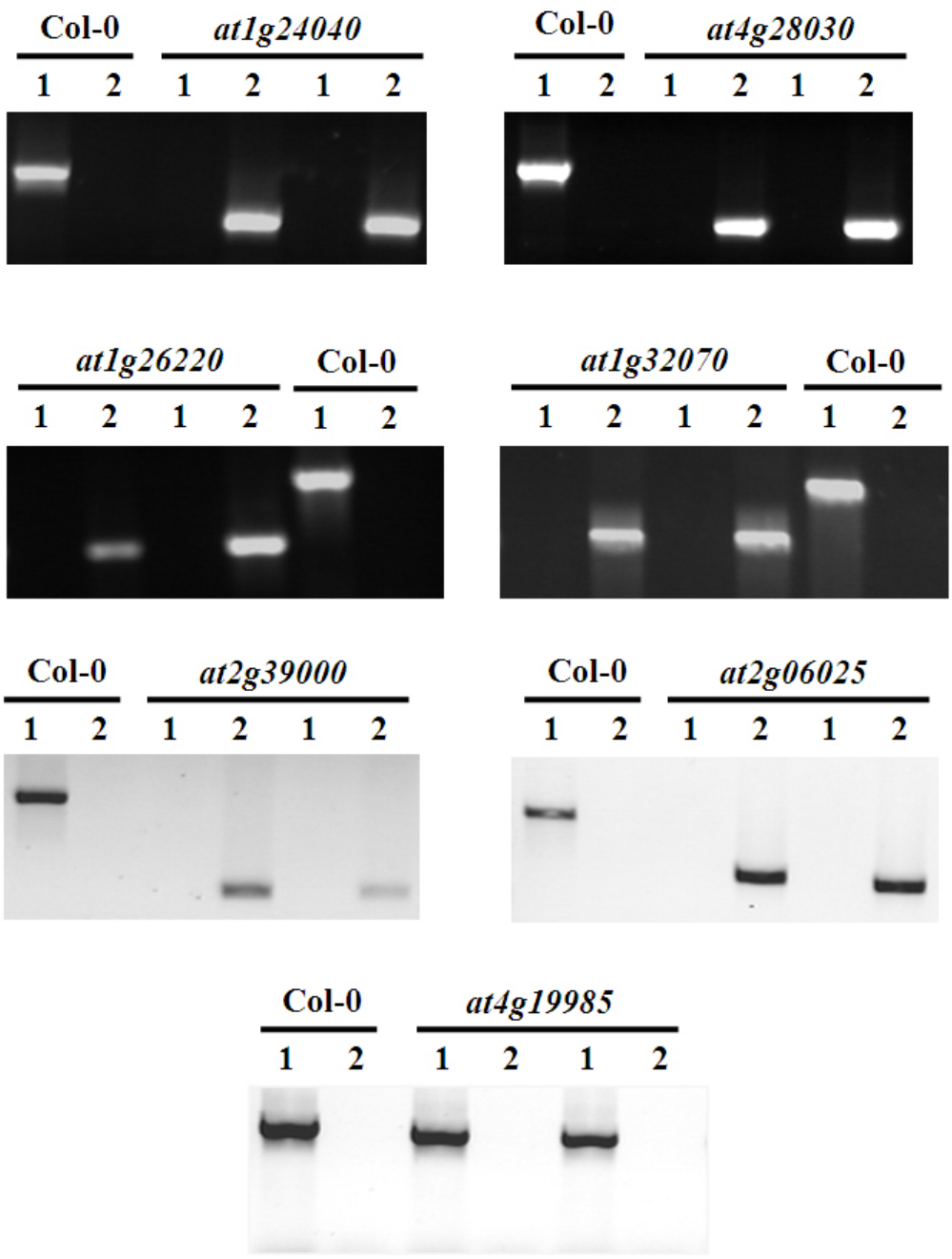


Figure S2. PCR genotyping of different plastidic NAT T-DNA insertion mutants.

PCR genotyping was run with two pairs of primers, gene specific primers (1) and T-DNA insertion specific primers (2).

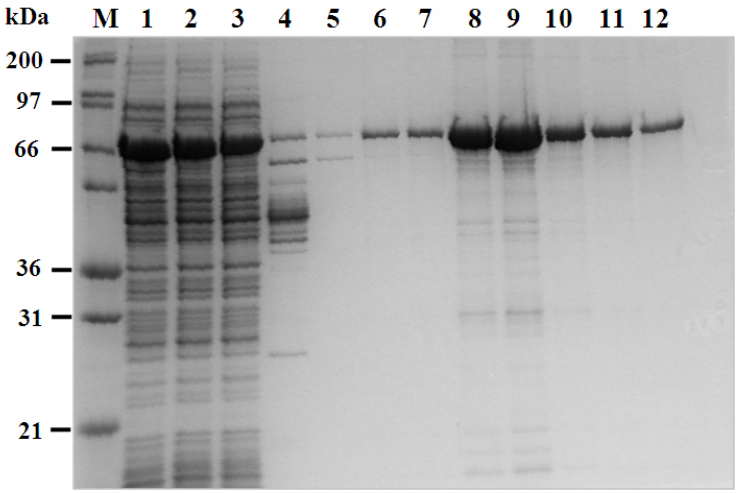


Figure S3. Purification of MBP-pNaa10p.

The expression of recombinant MBP-pNaa10p was induced by 1mM IPTG for 5 hours at 37^{o}C and purified using HiTrap Chelating HP column. Proteins from the different fractions of purification procedure were separated using SDS-PAGE and visualized by Coomassie staning. (M) Mark12 protein ladder, (1) Crude extract – 10 µg, (2) Flow through extract – 10 µg, (3) Wash fraction 1 – 10 µg, (4-6) Wash fractions 5,10,15 – maximum 2 µg of protein was loaded, (7-12) Elution fractions 1-6 – maximum 2 µg of protein was loaded.

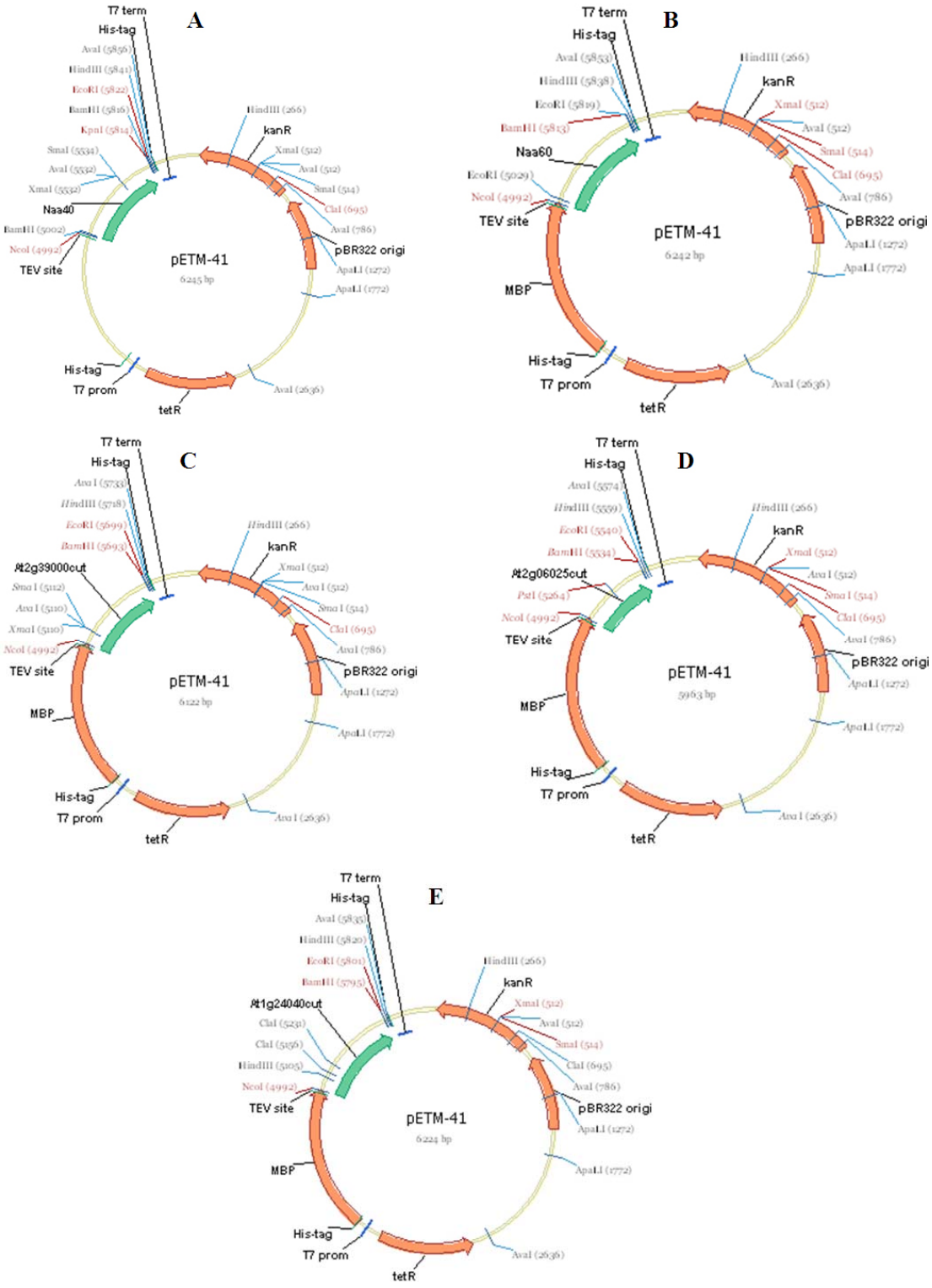


Figure S4. Vector maps of constructs that were created in this study using original vector pETM-41. Vector maps were created by Vector NTI 9. (A) pETM41-Naa40, (B) pETM41-Naa60, (C) pETM41-pNaa10, (D) pETM41-pNaa20, (E) pETM41-pNaa30.