

"It always seems impossible until it's done."

- Nelson Mandela -

Promotor: **Prof. dr. Els J.M. Van Damme**
Laboratory of Biochemistry and Glycobiology
Department of Molecular Biotechnology
Faculty of Bioscience Engineering
Ghent University

Dean: **Prof. dr. ir. Marc Van Meirvenne**

Rector: **Prof. dr. Anne De Paepe**

**Nictaba lectin homologs from
Arabidopsis thaliana and their putative
role in plant stress responses**

Lore Eggermont

**Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology,
Faculty of Bioscience Engineering, Ghent University**

Thesis submitted in fulfillment of the requirements for the degree of Doctor (PhD) in Applied
Biological Sciences: Cell and Gene Biotechnology

Dutch translation of the title:

De vermoedelijke rol van Nictaba lectine homologen van *Arabidopsis thaliana* in de reactie op plantenstress

Cover:

Arabidopsis thaliana, adapted from the picture from Dr. Barry Causier, Centre for Plant Sciences, University of Leeds.

Cite as:

Eggermont L (2017). Nictaba lectin homologs from *Arabidopsis thaliana* and their putative role in plant stress responses. PhD thesis, Ghent University, Belgium.

ISBN number: 978-94-6357-022-0

The author and the promotor give the authorization to consult and to copy parts of this work for personal use only. Any other use is limited by the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

The promotor,

Prof. dr. Els J.M. Van Damme

The author,

Lore Eggermont

Members of the examination committee

Prof. dr. Els J.M. Van Damme (promotor)

Laboratory of Biochemistry and Glycobiology
Department of Molecular Biotechnology
Faculty of Bioscience Engineering, Ghent University

Prof. dr. ir. Mieke Uyttendaele (chair)

Laboratory for Food Microbiology and Food Preservation
Department of Food Safety and Food Quality
Faculty of Bioscience Engineering, Ghent University

Prof. dr. ir. Kris Audenaert

Laboratory of Plant Production
Department of Applied Biosciences
Faculty of Bioscience Engineering, Ghent University

Prof. dr. ir. Tina Kyndt

Laboratory of Epigenetics and Defence
Department of Molecular Biotechnology
Faculty of Bioscience Engineering, Ghent University

Prof. dr. Filip Vandenbussche

Laboratory of Functional Plant Biology
Department of Biology
Faculty of Sciences, Ghent University

Prof. dr. Filip Rolland

Laboratory of Molecular Plant Biology
Department of Biology
Faculty of Science, Katholieke Universiteit Leuven

Table of contents

Members of the examination committee	i
Table of contents	iii
List of abbreviations	ix
Introduction	1
1.1 Plant stress in <i>Arabidopsis thaliana</i>	3
1.1.1 <i>A. thaliana</i> as model organism	3
1.1.2 Plant stress	4
1.1.2.1 Abiotic stress	4
1.1.2.1.1. Osmotic stress.....	5
1.1.2.1.2. Heat stress	10
1.1.2.2 Biotic stress.....	11
1.1.2.2.1. Plant pathogens	14
1.2 Lectins.....	19
1.3 Nictaba and homologs in <i>A. thaliana</i>	20
1.3.1 Nictaba.....	20
1.3.1 Nictaba homologs in <i>A. thaliana</i> (ArathNictabas)	20
Scope	23
Genome-wide screening for lectin motifs in <i>Arabidopsis thaliana</i>	25
2.1 Abstract	27
2.2 Introduction.....	27
2.3 Materials and methods	28
2.3.1 Identification of the putative lectin genes in <i>A. thaliana</i>	28
2.3.2 Annotating the protein domains of putative lectins	29
2.3.3 Determining signal sequences and/or transmembrane domains.....	29
2.3.4 Mapping the putative lectin genes on the chromosomes of Arabidopsis and analysis of tandem duplications.....	30
2.3.5 Phylogenetic analysis.....	30
2.3.6 Analysis of amino acids responsible for carbohydrate binding.....	30
2.4 Results and discussion.....	31
2.4.1 Identification and distribution of the genes with a lectin domain in <i>A. thaliana</i>	31
2.4.2 Domain architecture and importance of the putative lectins in <i>A. thaliana</i>	32

2.4.2.1	CRA homologs.....	33
2.4.2.2	EUL homologs	34
2.4.2.3	GNA homologs.....	34
2.4.2.4	Hevein homologs.....	36
2.4.2.5	Jacalin homologs	37
2.4.2.6	Legume lectin homologs.....	38
2.4.2.7	LysM homologs.....	39
2.4.2.8	Nictaba homologs.....	41
2.4.2.9	Ricin B homologs	42
2.4.3	Phylogenetic analysis and analysis of the carbohydrate binding site.	42
2.4.3.1	Evolutionary relationships of the jacalin homologs from <i>A. thaliana</i>	43
2.4.3.2	Conserved amino acids in the carbohydrate binding site of jacalin homologs from <i>Arabidopsis</i>	45
2.4.3.3	Phylogenetic analysis of the LysM domain family in <i>A. thaliana</i>	46
2.4.3.4	Conservation of the carbohydrate binding site in the LysM homologs from <i>Arabidopsis</i>	48
2.4.3.5	Phylogenetic analysis of the Nictaba family in <i>A. thaliana</i>	49
2.4.3.6	Conserved amino acids in the carbohydrate binding site of Nictaba homologs from <i>Arabidopsis</i>	51
2.5	Conclusions.....	52
	Comparative analysis of the Nictaba homologs in <i>Arabidopsis thaliana</i>	53
3.1	Abstract	55
3.2	Introduction.....	55
3.3	Materials and methods	56
3.3.1	Plant material and growth conditions.....	56
3.3.2	<i>In silico</i> tools	57
3.3.3	Expression in normal growth conditions during plant development.....	57
3.3.4	Hormone and abiotic stress treatments	58
3.3.5	Biotic stress treatments.....	58
3.3.6	RNA extraction, cDNA synthesis and RT-PCR analysis.....	60
3.3.7	Construction of EGFP fusion constructs	60
3.3.8	Transient transformation of <i>N. benthamiana</i> leaves	61
3.3.9	Stable transformation of <i>A. thaliana</i> plants.....	62
3.3.10	Confocal microscopy and image analysis	62

3.3.11	qRT-PCR analysis.....	62
3.4	Results	63
3.4.1	Nictaba homologs in <i>A. thaliana</i>	63
3.4.2	Sequence similarity between the Nictaba homologs from <i>A. thaliana</i> and Nictaba	65
3.4.3	The Nictaba homologs from Arabidopsis show a nucleocytoplasmic localization.....	66
3.4.4	<i>In silico</i> expression analysis of the <i>Nictaba</i> homologs from <i>A. thaliana</i>	68
3.4.5	Expression of the <i>ArathNictaba</i> genes during development of WT <i>A. thaliana</i> plants	71
3.4.6	<i>ArathNictaba</i> expression is stress-inducible	73
3.4.6.1	<i>AN3</i> , <i>AN4</i> and <i>AN5</i> are differentially expressed in response to hormone treatments.....	73
3.4.6.2	The expression of the <i>ArathNictabas</i> showed dissimilar patterns after abiotic stress treatments.....	75
3.4.6.3	Expression of the <i>ArathNictabas</i> after different biotic stresses	76
3.5	Discussion	78
	Functional study of Nictaba homologs in <i>Arabidopsis thaliana</i>.....	83
4.1	Abstract	85
4.2	Introduction.....	85
4.3	Materials and methods	86
4.3.1	Plant material and growth conditions.....	86
4.3.2	Construction of <i>ArathNictaba</i> overexpression constructs	87
4.3.3	Stable transformation of <i>A. thaliana</i> plants.....	87
4.3.4	RNA extraction, cDNA synthesis, RT-PCR and qRT-PCR analysis	87
4.3.5	Checking T-DNA insertion lines for <i>AN4</i> and <i>AN5</i>	87
4.3.6	Germination assays and salt stress tolerance	88
4.3.7	Biotic stress susceptibility experiment.....	88
4.3.7.1	Quantification of leaf damage.....	89
4.3.7.2	Visualisation and quantification of cell death	89
4.3.7.3	Determination of <i>P. syringae</i> biomass	89
4.4	Results	90
4.4.1	Selection of T-DNA insertion and overexpression lines	90
4.4.1.1	Available <i>AN4</i> and <i>AN5</i> T-DNA insertion lines contain no T-DNA insertion.....	90
4.4.1.2	Overexpression lines for <i>AN3</i> , <i>AN4</i> and <i>AN5</i>	91
4.4.2	Germination assay of different transgenic lines compared to WT <i>A. thaliana</i> under normal growth and salt stress conditions	92

4.4.3	<i>ArathNictaba</i> overexpression lines show less disease symptoms and bacterial growth after <i>P. syringae</i> infection	93
4.5	Discussion	97
Recombinant expression of the Nictaba homologs and interaction study		99
5.1	Abstract	101
5.2	Introduction.....	101
5.3	Materials and methods	103
5.3.1	Plant material and growth conditions.....	103
5.3.2	Construction of <i>ArathNictaba</i> His6-tagged constructs	103
5.3.3	Transformation of <i>E. coli</i> and expression analysis	105
5.3.4	Optimization of recombinant protein expression in <i>E. coli</i>	106
5.3.5	Protein purification using column chromatography	106
5.3.6	SDS-PAGE and Western blot.....	106
5.3.7	Agglutination assay.....	107
5.3.8	Affinity chromatography with carbohydrates and glycoproteins	107
5.3.9	Pull-down analysis.....	107
5.3.10	Molecular modelling.....	109
5.4	Results	110
5.4.1	Recombinant expression and purification of AN4 in <i>E. coli</i>	110
5.4.2	Pull-down analysis to search for interacting partners of AN4.....	114
5.4.3	Molecular modelling of AN4.....	117
5.5	Discussion	120
General discussion and future perspectives		125
6.1	<i>ArathNictabas</i> are part of the large family of plant lectins and are widespread in angiosperms	127
6.2	<i>ArathNictabas</i> are low expressed in the nucleus and the cytoplasm	129
6.3	The expression of <i>ArathNictabas</i> is stress-inducible	132
6.3.1	Abiotic stress	132
6.3.2	Biotic stress.....	134
6.3.3	Cross-talk between abiotic and biotic stress.....	136
6.4	AN4 interacts with two plant defence-involved enzymes	136
Summary.....		141
Samenvatting		145

Supplementary data	151
Supplementary tables.....	153
Supplementary figures	174
References	177
Curriculum vitae	201

List of abbreviations

aa	amino acids
ABA	abscisic acid
ABAR	ABA-receptor
ABRE	ABA-responsive element
AGI	Arabidopsis Gene Initiative
AI1	avirulence induced gene 1
AN3	ArathNictaba3
AN4	ArathNictaba4
AN4-HIS	C-terminally His6-tagged AN4
AN5	ArathNictaba5
AP2	apetala2
ArathNictaba	<i>A. thaliana</i> Nictaba homolog
ATP	adenosine-5'-triphosphate
Avr	avirulence
BAK1	BRI1-associated receptor kinase 1
BiFC	bimolecular fluorescence complementation
BIK1	botrytis-induced kinase 1
BLAST	Basic Local Alignment Search Tool
BLASTn	nucleotide BLAST
BRCA2A	breast cancer2A
BSK1	brassinosteroid-signalling kinase 1
BY-2	bright yellow-2
CaM	calmodulin
CBF	C-repeat binding factor
CBL	calcineurin B-like protein
CDPK	calcium dependent protein kinase
CERK1	chitin elicitor receptor kinase 1
CHLH	chelataase H subunit
CID	chitinase insertion domain
Col	Columbia

CRA	class V chitinase-related agglutinin
CRT	cold-responsive element or C-repeat
Cul3	Cullin 3 E3 ligase
DAG	diacylglycerol
DAMP	damage-associated molecular pattern
dpi	days post infection/infestation
DRE	dehydration-responsive element
DREBs	dehydration-responsive element binding
EFR	EF-Tu receptor
EF-Tu	elongation factor Tu
EGFP	enhanced green fluorescent protein
EST	expressed sequence tag
ET	ethylene
ETI	effector-triggered immunity
ETS	effector-triggered susceptibility
EUL	<i>Euonymus europaeus</i> lectin
Flg22	22-amino-acid conserved peptide of bacterial flagellin
FLS2	flagellin sensing 2
GDP	guanosine-5'-diphosphate
GH	glycosyl hydrolase
GlcNAc	<i>N</i> -acetyl-D-glucosamine
GNA	<i>Galanthus nivalis</i> agglutinin
GO	Gene Ontology
GPA1	G protein alpha subunit 1
GSNO	S-nitrosoglutathione
GTG	G protein-coupled receptor-type G protein
GTP	guanosine-5'-triphosphate
GTPase	guanosine-5'-triphosphatase
HIS-AN4	N-terminally His6-tagged AN4
HR	hypersensitive response
HSE	heat shock cis-regulatory promoter element
Hsf	heat stress factor

Hsp	heat shock protein
ICS/SID2	isochorismate synthase
InsP	inositol phosphates
IP3	inositol 1,4,5-trisphosphate
IZ	imidazole
JA	jasmonic acid
JRL	jacalin-related lectin
Ler	Landsberg
LRR	leucine-rich repeat
LYK5	LysM-containing receptor-like kinase 5
LYM	LysM domain-containing GPI-anchored protein
LysM	Lysin Motif
MAMP	microbial-associated molecular pattern
MAPK	mitogen-activated protein kinase
MBP	myrosinase binding protein
MeJA	methyl jasmonate
M-MLV	moloney murine leukemia virus
MOS	modifier of <i>snc1</i>
MS	mass spectrometry
MS	Murashige and Skoog
NCED	9-cis-eposycarotenoid dioxygenase
Nictaba	<i>Nicotiana tabacum</i> agglutinin
NIMIN	NIM1-interacting protein
NLR	nucleotide-binding leucine-rich repeat
NLS	nuclear localization signal
NO	nitric oxide
NPR1	non-expressor of PR genes 1
P	phosphorylation
PAGE	polyacrylamide gel electrophoresis
PAL	phenylalanine ammonia lyase
PAMP	pathogen-associated molecular pattern
PB	phosphate buffer

PBL27	PBS1-like kinase 27
PBP1	PYK10 binding protein 1
PBS1	AvrPphB susceptible 1
PCRK1	pattern-triggered immunity compromised receptor-like cytoplasmic kinase 1
PEPR	Pep receptor
Phe	phenylalanine
PIP2	phosphatidylinositol 4,5-bisphosphate
PLC	phospholipase C
PP2	phloem protein 2
PP2A	protein phosphatase 2A
PP2C	type 2C protein phosphatase
PPV	plum pox virus
PR	pathogenesis related
PRR	pattern recognition receptor
PTI	PAMP-triggered immunity
PYL	pyrabactin resistance-like
PYR	pyrabactin resistance
qRT-PCR	quantitative reverse transcriptase-polymerase chain reaction
R	resistance
RAD51(D)	ras associated with diabetes51(D)
RBOH	respiratory burst oxidative homolog
RCAR	regulatory component of ABA receptors
RCLK	receptor-like cytoplasmic kinase
RLK	receptor-like kinase
RLP	receptor-like protein
ROS	reactive oxygen species
SA	salicylic acid
SAR	systemic acquired resistance
SDS	sodium dodecyl sulphate
SERK	somatic embryogenesis receptor kinase
SMART	Simple Modular Architecture Research Tool

SNI1	suppressor of <i>npr1</i> inducible1
SnRK2	SNF1-related protein kinase
SOS	salt overly sensitive
SRK	S-locus receptor kinase
SSN2	suppressor of <i>sni1 2</i>
T3SS/TTSS	type III secretion system
TAIR10	The Arabidopsis Information Resource 10
tBLASTn	translated nucleotide BLAST
T-DNA	transferred DNA
TEV	tobacco etch potyvirus
TF	transcription factor
TIR	Toll/Interleukin-1 receptor
TRXs	thioredoxins
Ub	ubiquitinylation
VBF	VIP1 binding F-box protein
WT	wild type

Chapter 1

Introduction

1.1 Plant stress in *Arabidopsis thaliana*

1.1.1 *A. thaliana* as model organism

A. thaliana is the most widely-studied plant and serves as a model organism to understand the complex processes required for plant growth and development (Rhee *et al.*, 2003). Furthermore, many aspects of adaptation of plants to adverse conditions and diseases are studied in *A. thaliana* (Delseny and Pelletier, 2001). For almost every aspect of life, humans depend on plants and as a result of an increasing world population and climate instability, knowledge of plant processes will be crucial to meet the exponentially increasing requirements for food and fuel supplies (Lavagi *et al.*, 2012). Although research on *A. thaliana* can facilitate the identification of related genes of importance in crop plants and as such can help to solve problems related to agriculture, energy, industry, human health and the environment, it should be noted that the extrapolation from *A. thaliana* to crops is not always possible (Meinke *et al.*, 1998; Sivasubramanian *et al.*, 2015).

A. thaliana is a small flowering plant which is part of the Brassicaceae or mustard family. One of the advantages to work with this model organism is the fully sequenced and annotated small genome. Moreover, the small size of the plant, the rapid generation time, the ability of self-pollination, the large seed set and the establishment of transformation protocols are other advantages of this model organism. All these favourable characteristics make *A. thaliana* into the reference plant for plant biology (Meinke *et al.*, 1998; Delseny and Pelletier, 2001; Lavagi *et al.*, 2012). Many ecotypes have been gathered from natural populations and are available, but the Columbia (Col) and Landsberg *erecta* (Ler) ecotypes are the conventional standards for research (Meinke *et al.*, 1998).

The genome of *A. thaliana* was sequenced by the Arabidopsis Genome Initiative in 2000 (The Arabidopsis Genome Initiative, 2000). The genome of 135 megabases is organized into five chromosomes (TAIR10¹;Meinke *et al.*, 1998). Out of a total of 33,602 genes; 27,416 genes are coding for proteins; 4,827 are pseudogenes or transposable element genes and 1,359 sequences represent non coding RNAs (TAIR10¹). As these numbers indicate, the genome of *A. thaliana* is highly enriched for coding sequences. On average every five kilobases one protein coding gene is present in the Arabidopsis genome. Moreover, half of these protein coding genes are found to be closely related to genes from other organisms ranging from bacteria to humans (Meinke *et al.*, 1998). According to The Arabidopsis Information Resource 10 (TAIR10) genome release, at least one Gene Ontology (GO) annotation is available for 77 % of all Arabidopsis genes (Lamesch *et al.*, 2012). A large number of genes are part of multigene families, not so surprisingly knowing that almost 60 % of the *A. thaliana* genome encompasses duplicated regions (The Arabidopsis Genome Initiative,

¹ https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation

2000; Delseny and Pelletier, 2001). These gene duplicates arose by segmental, whole genome and tandem duplication events. Three whole genome duplication events occurred in *A. thaliana* after its separation from rice. The impact of these whole genome duplications is large because all genes were duplicated at once. In contrast, each tandem duplication only involves a small number of genes. However, tandem duplications contributed significantly to the expansion of gene families, approximately 17 % of all Arabidopsis genes are part of tandem duplications (The Arabidopsis Genome Initiative, 2000; Hanada *et al.*, 2008).

1.1.2 Plant stress

Plant stress groups all unfavourable environmental conditions which create potentially damaging physiological changes within plants. Plants are often exposed to a combination of different stresses which have negative effects on plant growth and development (Atkinson and Urwin, 2012; Osakabe *et al.*, 2013). The combination of these stresses not only threatens important crops but also plants in natural environments that are part of the ecosystem (Gassmann *et al.*, 2016). Since plants are sessile, they evolved complex adaptive and defence mechanisms in response to those environmental stresses. These mechanisms are activated in the tissues exposed to stress but also in distal portions of the plant, not directly exposed to stress (Osakabe *et al.*, 2013; Baxter *et al.*, 2014).

Plant stress is generally classified into abiotic and biotic stresses. Abiotic stresses such as heat, cold, drought, salt, water, light, ... have a large impact on world agriculture. They can reduce the average yields of major crop plants by more than 50 %. Biotic stresses are caused by fungi, bacteria, viruses, nematodes and herbivorous insects. Each stress, abiotic or biotic, evokes a complex cellular and molecular response to prevent damage and ensure survival of the plant. These responses cost energy and consequently diminish the growth and the yield of the plants (Atkinson and Urwin, 2012; Huber and Bauerle, 2016).

1.1.2.1 Abiotic stress

Abiotic stresses, such as drought, salinity, extreme temperatures, chemical toxicity and oxidative stress cause serious crop losses in agriculture throughout the world, the average yields for most major crop plants are diminished by more than 50 % (Wang *et al.*, 2003; Sewelam *et al.*, 2014). Moreover in future, the timing of abiotic stress will be less predictable in addition to more severe abiotic stress. Additionally, the influence of multiple abiotic stresses is expected to rise significantly with climate change. Because of a rapidly expanding world population, the need for food rises. To secure our future food supply, crop plants adapted to survive the environmental stress will be necessary (Kazan, 2015; Le Gall *et al.*, 2015).

Because of the vast amount of data related to implications of abiotic stress in *A. thaliana* this chapter will focus on some major threats, in particular osmotic and heat stress, stress treatments that will also be investigated in chapter 3. Drought, salinity and extreme temperatures are often interconnected and activate similar signalling pathways and cellular

responses. Drought and high salinity stress appear primarily as osmotic stress and as such both result in the disruption of homeostasis and ion distribution in the cell. On the other hand, all three stresses cause oxidative stress, which results in the denaturation of important biomolecules. Typical cellular responses on this osmotic and oxidative stress are the production of stress proteins and anti-oxidants, and the accumulation of compatible solutes (Wang *et al.*, 2003).

1.1.2.1.1. Osmotic stress

In nature, high salinity and drought are the major causes of osmotic stress (Xiong and Zhu, 2002). Drought and salt stress affect more than 10 % of the arable land, resulting in crop losses with a serious economic impact and which are predicted to increase with global climate change (Roychoudhury *et al.*, 2013). Global climate changes are resulting in increased temperature and atmospheric CO₂ levels as well as changing rainfall patterns. As a consequence, periods of inadequate rainfall will appear more frequently, leading to more drought stress (Bhargava and Sawant, 2013).

Osmotic stress results in disruption of homeostasis and ion distribution in the cell (Wang *et al.*, 2003). Upon osmotic stress, plants react with a wide range of responses at molecular, cellular and morphological levels (Xiong and Zhu, 2002; Wang *et al.*, 2003). Examples of responses at morphological level are the inhibition of shoot growth and the enhancement of root growth. Examples of cellular responses are the adjustment of ion and water transport, and metabolic changes (e.g. synthesis of compatible solutes or osmoprotectants). All these cellular and morphological responses are regulated at the molecular level by inducing the expression of stress-responsive genes (Xiong and Zhu, 2002; Golldack *et al.*, 2014).

The phytohormone abscisic acid (ABA) plays a crucial role during osmotic stress. Many drought and high salinity-responsive genes, respond to ABA. However, next to this ABA-dependent pathway, some drought-responsive genes are regulated independent of ABA (Roychoudhury *et al.*, 2013). The ABA-dependent pathway is described in the 'Hormonal signalling' section. ABA-independent regulation uses the cis-acting promoter elements, named dehydration-responsive element (DRE) and cold-responsive element or C-repeat (CRT). Both contain the core CCGAC sequence which suggest a cross-talk between cold and drought stimuli. Transcription factors (TFs) belonging to the apetala2 (AP2)-type family, called dehydration-responsive element binding (DREBs) bind DREs in the promoters of stress-responsive genes and as such activate these genes. TFs that can bind CRT cis-acting elements are called C-repeat binding factors or CBFs (Wang *et al.*, 2003; Bhargava and Sawant, 2013; Roychoudhury *et al.*, 2013). Calcium is a secondary messenger in the cross-talk between the ABA-dependent and ABA-independent pathway (Vishwakarma *et al.*, 2017). ABA, drought stress, cold stress and high salt stress all induce high intracellular calcium levels in plant cells, which is sensed by different calcium sensors (Figure 1.2) (Mahajan and Tuteja, 2005).

Hormonal signalling

ABA, a phytohormone, is a key regulator in the plant stress response to drought and salinity, and has a crucial function as a growth inhibitor (Golldack *et al.*, 2014). ABA plays also an important role toward a wide range of other stresses like heavy metal, heat, cold, radiation stress and some biotic stresses (Vishwakarma *et al.*, 2017). Moreover, ABA functions in seed germination, seed dormancy, closure of stomata, carbohydrate and lipid metabolism (Golldack *et al.*, 2014; Vishwakarma *et al.*, 2017).

Upon dehydration and high salinity, ABA *de novo* synthesis takes place and during rehydration, its degradation occurs in plant roots and terminal buds at the top of the plant. ABA biosynthesis starts from β -carotene with the involvement of various enzymes. The increase in the endogenous ABA levels is due to stress-related induction of genes that code for the ABA biosynthesis enzymes (Roychoudhury *et al.*, 2013; Vishwakarma *et al.*, 2017). One of these biosynthesis enzymes is 9-cis-epoxycarotenoid dioxygenase (NCED) (Vishwakarma *et al.*, 2017). NCED3 from *A. thaliana* is most strongly induced by drought and salt stress (Roychoudhury *et al.*, 2013). Overexpression of NCED3 in *A. thaliana* revealed an improvement of shoot growth under drought stress and its knockout mutant showed a dehydration sensitive phenotype (Iuchi *et al.*, 2001; Roychoudhury *et al.*, 2013).

ABA perception occurs via multiple cellular receptors that function in different subcellular compartments (Figure 1.1)(Shan *et al.*, 2012). The first class are the nucleocytoplasmic receptors pyrabactin resistance/pyrabactin resistance-like/regulatory component of ABA receptors (PYR/PYL/RCAR) which can bind intracellular ABA and as such inhibit type 2C protein phosphatases (PP2Cs) such as ABI1 and ABI2. Due to the inactivation of these PP2Cs, active SNF1-related protein kinases (SnRK2s) accumulate and phosphorylate ABA-responsive TFs which in turn activate ABA-responsive genes (Golldack *et al.*, 2014). These receptors mediate the ABA-regulated seed germination, growth and guard cell movement (Shan *et al.*, 2012). The second class are plasma membrane-localized ABA receptors which are G protein-coupled receptor-type G proteins (GTGs). ABA can be bound by GTG1/GTG2 which contain nucleotide-binding and guanosine-5'-triphosphatase (GTPase)-activating domains. The binding of ABA by GTGs is enhanced by guanosine-5'-diphosphate (GDP). The G protein alpha subunit 1 (GPA1) interacts with GTG1 and GTG2, and represses their GTPase activity. ABA binding is not inhibited since GDP only enhances this binding. These receptors have been shown to function as voltage-dependent anion channels and as such play a role in growth and development. A third kind of ABA receptor is localized in the chloroplast namely the chelatase H subunit (CHLH) of Mg-chelatase. CHLH transfers the signal to the nucleus via binding of its C-terminus to the transcription repressors WRKY40, WRKY18 and WRKY60 in the cytoplasm. This receptor mediates seed germination, seedling growth and guard cell movement (Shan *et al.*, 2012; Golldack *et al.*, 2014).

Note that there is some controversy related to the chloroplast CHLH and the plasma membrane GTG ABA receptors (Fan *et al.*, 2016; Verslues, 2016). Tsuzuki *et al.* (2011)

demonstrated that recombinant CHLH did not bind ABA, while recombinant PYR showed specific ABA binding. In contrast, ABA binding was shown for recombinant GTGs (Kharenko *et al.*, 2013).

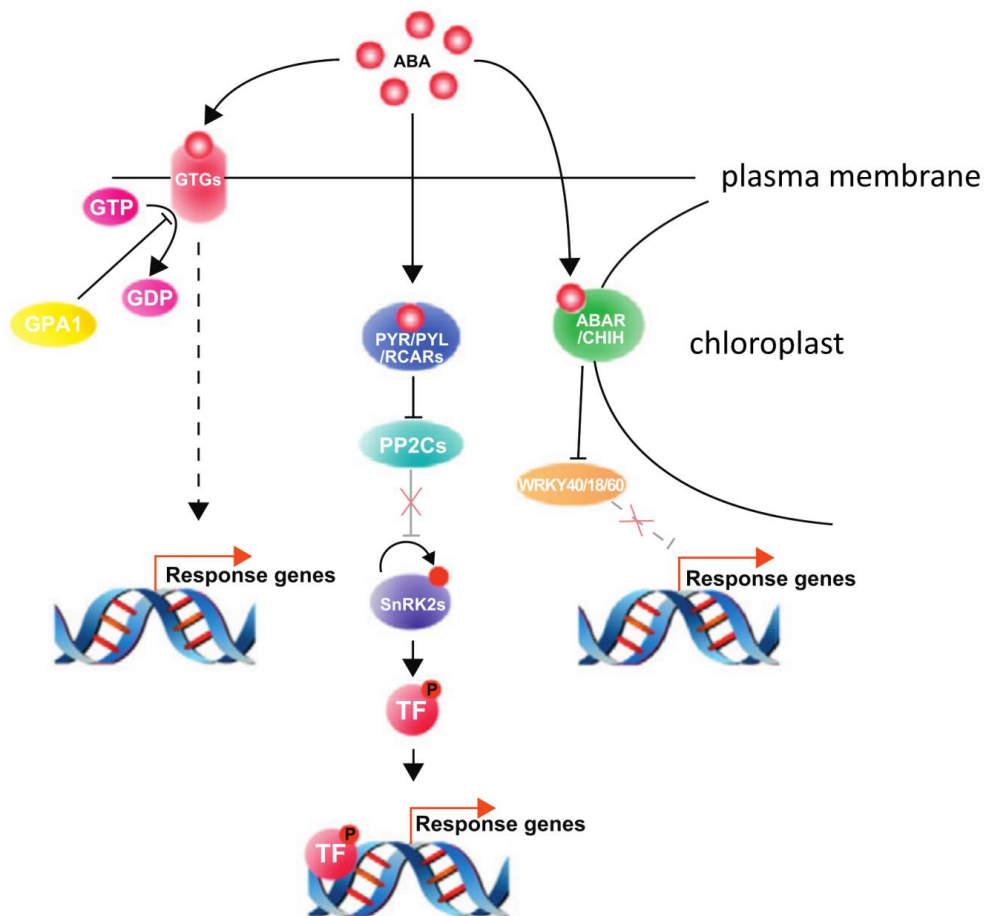


Figure 1.1 ABA perception and signalling (Adapted from Shan *et al.*, 2012). Multiple cellular receptors function in different subcellular compartments for ABA perception. The PYR/PYL/RCAR receptors are nucleocytoplasmic ABA receptors and inhibit PP2Cs upon ABA binding. Inhibition of PP2Cs leads to the activation of SnRK2s and as such phosphorylation of ABA-responsive TFs, which in turn activate ABA-responsive genes. The GTG ABA receptors are localized on the plasma membrane and GDP enhances the binding of ABA. After binding of ABA, the signal is transferred to the nucleus. The CHLH/ABAR ABA receptors are localized in the chloroplast membrane. After binding of ABA, CHLH/ABAR inhibit the transcription repressors WRKY40, WRKY18 and WRKY60 and as such ABA-responsive genes are activated. PYR: pyrabactin resistance, PYL: pyrabactin resistance-like, RCAR: regulatory component of ABA receptors, PP2C: type 2C protein phosphatase, SnRK2: SNF1-related protein kinase, GTG: G protein-coupled receptor-type G protein, GTP: guanosine-5'-triphosphate, GDP: guanosine-5'-diphosphate, GPA1: G protein alpha subunit 1, CHLH: chelatase H subunit, ABAR: ABA-receptor, TF: transcription factor, P: phosphorylation.

Eventually ABA signalling controls the movement of stomata, tissue hydraulic conductivity, growth of root and shoot, and communication between root and shoot (Vishwakarma *et al.*, 2017). Roots sense changes in abiotic factors (e.g. drought). Consequently, ABA is synthesized in the roots and transported through the xylem to the shoots. ABA is one of the chemical signals for root-to-shoot communication, next to pH, cytokinins and ethylene

precursors. In the shoots, ABA is transported to the guard cells, where it regulates the closure of the stomata (Malladi and Burns, 2007; Zhang *et al.*, 2015). Moreover, Kuromori and Shinozaki (2010) reported two ATP-binding cassette (ABC) transporter genes in *Arabidopsis* which function as ABA transporters (export or import). The stomatal closure, initiated by ABA during drought stress, prevents more water loss via transpiration. Nitric oxide (NO) is an important compound in the ABA driven closure of stomata. This closure is mediated by guard cell depolarization and alterations of guard cell turgor and volume (Roychoudhury *et al.*, 2013). Also, it was shown that ABA mediates the production of reactive oxygen species (ROS), which can on their turn, activate defensive responses (Sakamoto *et al.*, 2008).

The promoters of the genes controlled by ABA contain ABA-responsive elements (ABREs) which have an ACGT core. Several TFs belonging to the basic leucine zipper (bZIP) and MYB family are ABRE-binding proteins (AREBs/ABFs). These TFs can induce ABA-responsive genes through binding of the ABREs in their promoter sequence (Wang *et al.*, 2003; Bhargava and Sawant, 2013; Roychoudhury *et al.*, 2013). The expression of these stress-responsive genes should ultimately lead to stress tolerance or resistance. Three major categories of stress-related genes are identified and include those involved in signalling cascades and transcriptional control, those with a function in protecting membranes and proteins, and those required for water/ion uptake and transport (Wang *et al.*, 2003).

Drought stress

Drought or dehydration is defined as an imbalance between soil water availability and evaporative demand. Plant growth and yield is reduced by this major environmental stress and the following physiological changes are generated: loss of turgor, osmotic adjustment and reduced leaf water potential. The low turgor pressure causes a reduction or cessation of growth by decreasing cell extensibility and cell expansion (Le Gall *et al.*, 2015).

Drought stress is sensed by a membrane-bound two-component histidine kinase (HK), an osmoreceptor. This receptor, ATHK1 in *A. thaliana*, in turn activates phospholipase C (PLC), which hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP₂) to inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol (DAG) (Figure 1.2). IP₃, as second messenger, releases calcium from internal stores (endoplasmic reticulum, vacuole and chloroplast) and these calcium molecules are sensed by Ca²⁺-sensors (e.g. calcineurin B-like protein (CBL) and calmodulin (CaM)). These sensors activate downstream protein kinases and phosphatases leading to the activation of TFs that bind DREs and induce the expression of several drought-responsive genes (Mahajan and Tuteja, 2005; Beck *et al.*, 2007; Tran *et al.*, 2007). This pathway is ABA-independent but ATHK1 also positively regulates stress responses through the ABA-dependent pathway (Tran *et al.*, 2007).

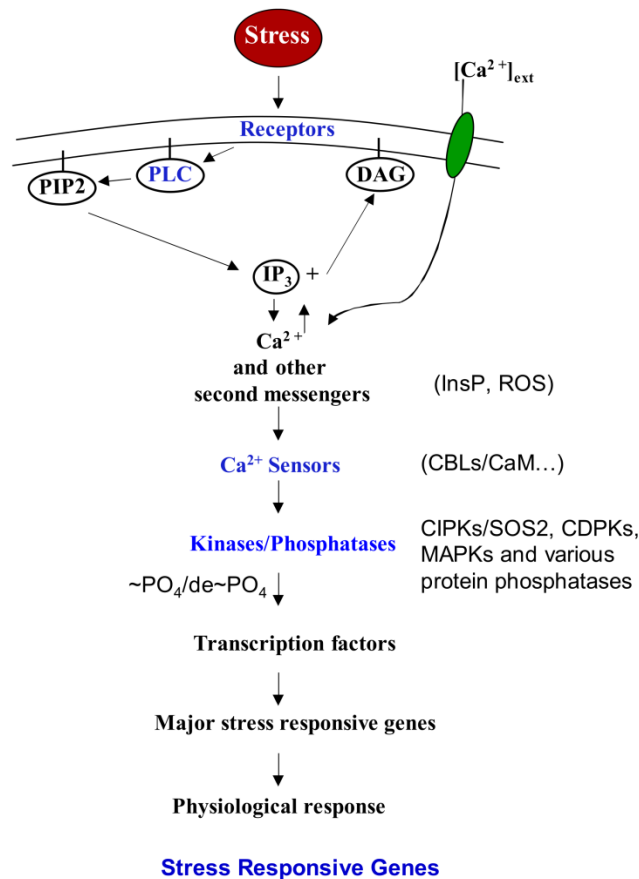


Figure 1.2 Drought perception and signalling (Adapted from Mahajan and Tuteja, 2005). Drought receptors activate PLC upon drought stress. PLC hydrolyzes PIP2 into IP3 and DAG. IP3 releases calcium from internal stores which is sensed by calcium sensors. Calcium sensors activate downstream kinases and phosphatases which in turn activates TFs that induce expression of several drought-responsive genes. PLC: phospholipase C, PIP2: phosphatidylinositol 4,5-bisphosphate, IP3: inositol 1,4,5-trisphosphate, DAG: diacylglycerol, InsP: inositol phosphates, CBL: calcineurin B-like protein, CaM: calmodulin.

Salt stress

Soil salinity is another major abiotic stress that affects plant growth and productivity. Up to 30 % of land loss is expected within the next 15 years caused by an increasing salinization of arable land (Le Gall *et al.*, 2015). Moreover, salinization may occur in more than 50 % of arable land by the year 2050 (Wang *et al.*, 2003). High salt concentrations in the soil can be caused via the deposition of oceanic salts by wind and rain and via the erosion of rocks that release soluble salts (Le Gall *et al.*, 2015). The major effect of salt stress to the plant is water deficit leading to growth inhibition and plant death during prolonged exposure (Zhu, 2007; Le Gall *et al.*, 2015). Most plants cannot tolerate salt stress and are called glycophytes. Glycophytes cannot tolerate soil salt concentrations of 40 mM and more. In contrast to that, halophytes naturally grow under high salinity conditions and can cope with salt soil concentrations higher than 200 mM (Zhu, 2007; Flowers and Colmer, 2008).

To recover ion homeostasis upon salt stress, plants developed the salt overly sensitive (SOS) pathway (Figure 1.3). The SOS pathway consists of three major proteins namely SOS1, SOS2 and SOS3. SOS1 encodes a plasma membrane Na^+/H^+ antiporter which is responsible for Na^+ -efflux, at the expense of H^+ . The sodium extrusion activity of SOS1 is dependent from SOS2 and SOS3. Salt stress is sensed by an unknown plasma membrane sensor causing cytoplasmic calcium perturbations. SOS2 is a serine/threonine protein kinase which can interact with SOS3. SOS3 encodes a Ca^{2+} -binding protein with a myristoylation sequence at the N-terminus of the protein. Upon calcium binding, the SOS3-SOS2 complex is activated and SOS2 can activate SOS1 by phosphorylation (Xiong and Zhu, 2002; Wang *et al.*, 2003; Mahajan and Tuteja, 2005; Zhu, 2007; Gupta and Huang, 2014). Shi *et al.* (2003) reported that overexpression of SOS1 results in an improved salt tolerance in transgenic *Arabidopsis*.

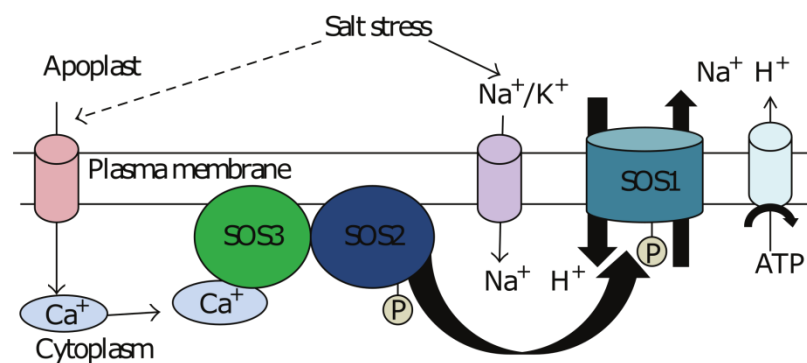


Figure 1.3 SOS pathway to recover the ion balance upon salt stress (Gupta and Huang, 2014). Unknown plasma membrane sensors cause higher levels of intracellular calcium upon salt stress. SOS3 senses calcium and activates SOS2, a serine/threonine protein kinase. The SOS3-SOS2 complex can activate SOS1 by phosphorylation. SOS1 is a plasma membrane Na^+/H^+ antiporter which is responsible for Na^+ -efflux. SOS: salt overly sensitive, P: phosphorylation, ATP: adenosine-5'-triphosphate.

1.1.2.1.2. Heat stress

In light of the global warming, heat stress will likely be an important abiotic stress to which plants will have to adapt. Heat stress is defined as an increasing temperature inducing irreversible damage to plant growth and development. Heat stress can lead to a serious yield reduction in many areas. Roots are more susceptible to heat stress than shoots. As a consequence high soil temperature is more harmful than high air temperature (Le Gall *et al.*, 2015).

One of the earliest responses upon heat stress is global inhibition of translation. Secondly, plants have to cope with the osmotic and oxidative stress caused by the initial heat stress. Heat stress factors (Hsfs) and heat shock proteins (Hsps) play a central role in the plant response upon heat stress. Hsfs are TFs that play an important role in the control of the expression of several heat responsive genes. Hsfs can bind to heat shock cis-regulatory promoter elements (HSEs) in the promoters of these heat responsive genes. Heat responsive genes consist of two major groups namely signalling components and functional genes.

Signalling components such as protein kinases and TFs, functional genes such as Hsps and catalase. Hsps and other chaperones play an important role to prevent protein unfolding and misfolding caused by heat stress. When this mechanism is overwhelmed, proteases degrade the unfolded and denaturated proteins (Wang *et al.*, 2003; Qu *et al.*, 2013; Echevarría-Zomeño *et al.*, 2016).

1.1.2.2 Biotic stress

Similar to abiotic stresses, biotic stresses affect the growth and the yield of major crop plants around the world. Biotic agents like bacteria, fungi, nematodes, aphids, ... can cause severe plant diseases and epidemics threatening crop yield and food security. Rapid spreading of these diseases over great distances occurs via wind, water, insects and humans (Dangl *et al.*, 2013). The knowledge of immune signalling upon perception of biotic stresses will provide the foundation to generate broad-spectrum disease resistant crop plants which will help to establish a more sustainable agriculture by replacing costly and unsustainable chemical controls (Dangl *et al.*, 2013; Couto and Zipfel, 2016).

Plants contain a well developed plant innate immune system which results in basal defence. If the invader is not able to circumvent this system, the basal defence leads to resistance of the plant (Dangl and Jones, 2001). Indeed, most plants are resistant to infection by most pathogens because of the plant innate immune system and are called non-hosts (Jones and Dangl, 2006). This immune system consist of two major mechanisms: pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) (Figure 1.4). PTI is triggered by molecular patterns common to many types of microorganisms and results in no symptoms or hypersensitive response (HR). ETI is triggered by the recognition of pathogen effectors and is often associated with HR, this local plant cell death can restrict the growth of the pathogen (Tsuda and Katagiri, 2010; Kushalappa *et al.*, 2016). Before invaders are confronted with PTI and ETI, they have to deal with several barriers at the plant surface such as wax layers, rigid cell walls, cuticular lipids, antimicrobial enzymes or secondary metabolites. Invaders first have to enter the host tissue through direct penetration of the plant surface, physical injuries or natural openings such as stomata (Muthamilarasan and Prasad, 2013).

PTI is also triggered by microbial-associated molecular patterns (MAMPs) and damage-associated molecular patterns (DAMPs), next to PAMPs (Muthamilarasan and Prasad, 2013). MAMPs are similar to PAMPs, all PAMPs are MAMPs, but not all MAMPs are PAMPs. PAMPs originate from pathogenic microorganisms, but MAMPs can also originate from non-pathogenic microorganisms (Boller and Felix, 2009). Several examples of MAMPs are bacterial flagellin, elongation factor Tu (EF-Tu), peptidoglycans, lipopolysaccharides and fungal chitin (Tsuda and Katagiri, 2010; Couto and Zipfel, 2016). DAMPs are generated as a consequence of damage to the structural barriers of plant tissues by lytic enzymes produced by the microbes. These DAMPs are localized in the apoplast, examples are cell wall fragments, cutin monomers and peptides (Muthamilarasan and Prasad, 2013).

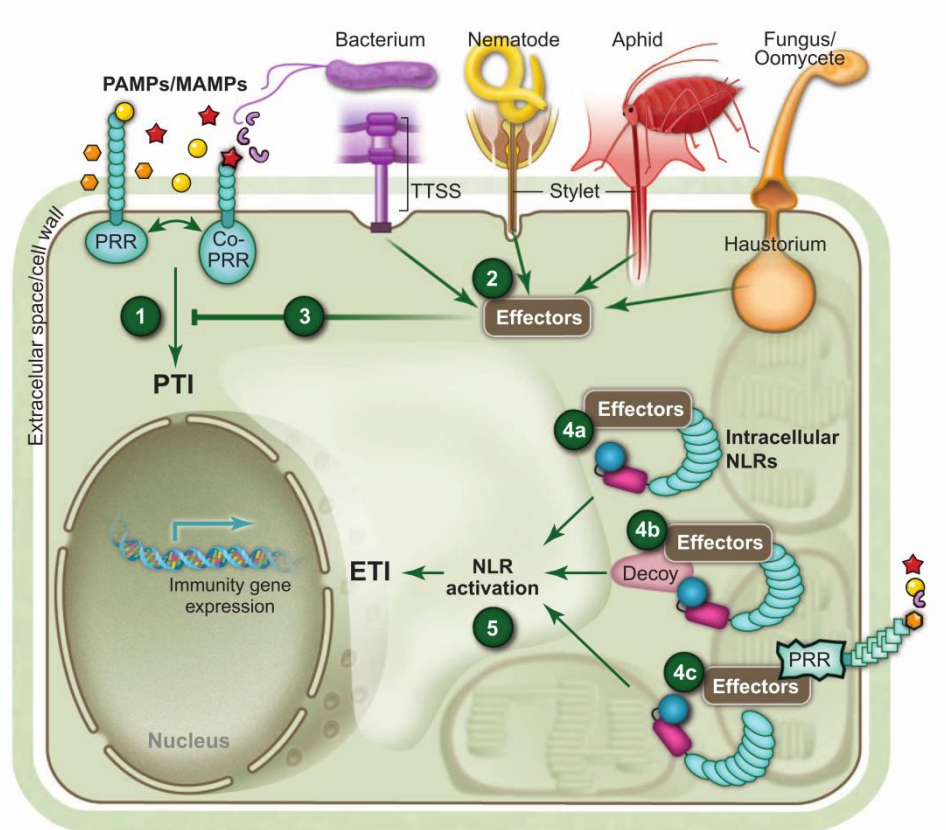


Figure 1.4 The plant immune system (Dangl *et al.*, 2013). 1) Recognition of MAMPs and DAMPs by PRRs to activate PTI. 2) Invaders secrete effector proteins. 3) Inhibition of PTI. 4) NLR receptors are activated through a) direct binding with effector proteins, b) binding with a decoy of the target from effectors or c) binding with a modified target from the effectors. PRR: pattern recognition receptor, PAMP: pathogen-associated molecular pattern, MAMP: microbial-associated molecular pattern, PTI: PAMP-triggered immunity, ETI: effector-triggered immunity, TTSS: type III secretion system (also T3SS), NLR: nucleotide-binding leucine-rich repeat.

The receptors on the plant cell surface that recognize the MAMPs and DAMPs are called pattern recognition receptors (PRRs) (Jones and Dangl, 2006). This PRR family resides in the plasma membrane and can be subdivided in two major groups of receptors. The first group includes receptor-like kinases (RLKs) which have an extracellular ligand-binding domain, a transmembrane domain and an intracellular serine/threonine kinase domain. The second group is called receptor-like proteins (RLPs) and are similar to RLKs except that they lack an intracellular kinase domain, to activate downstream signalling. As such these RLPs require interaction with adaptor molecules for signal transduction (Tsuda and Katagiri, 2010; Muthamilarasan and Prasad, 2013; Couto and Zipfel, 2016). The extracellular ligand-binding domains can vary and as such can bind different kinds of MAMPs. PRRs containing a leucine-rich repeat (LRR) prefer to bind proteins or peptides e.g. bacterial flagellin and EF-Tu. Other PRRs containing an extracellular Lysin Motif (LysM) domain, can bind carbohydrates such as bacterial peptidoglycans and fungal chitin whereas S-lectin domain containing PRRs bind extracellular ATP or lipopolysaccharides. All PRRs recruit regulatory receptor kinases upon ligand binding and further signalling proceeds through receptor-like cytoplasmic

kinases (RLCKs). These RLCKs are the link between MAMP or DAMP perception and downstream signalling. The affinity of RLCKs for different PRRs and the ability to activate different branches of the downstream signalling, varies among the large repertoire of RLCKs. In *Arabidopsis* 160 RLCKs are identified (Couto and Zipfel, 2016).

The recognition of MAMPs and DAMPs by PRRs is necessary to create PTI in plants (step 1, Figure 1.4) (Dangl *et al.*, 2013). Invaders can react to PTI by secreting effector proteins, encoded by avirulence (Avr) genes, in the plant cell which will inhibit PTI (step 2 and 3, Figure 1.4). The inhibition of PTI makes the plant susceptible and is called effector-triggered susceptibility (ETS) (Surico, 2013). Plants encode resistance (R) genes which encode nucleotide-binding leucine-rich repeat (NLR) receptors and react on these effectors resulting in ETI (step 4 and 5, Figure 1.4). Intracellular NLR receptors can sense the effector proteins in three different ways. NLR receptors can be activated through direct binding to the effectors (step 4a, Figure 1.4) or by a modified target from the effectors (or a decoy of the target) (step 4c and 4b, respectively, Figure 1.4) (Dangl *et al.*, 2013). There are 125 NLR receptors identified in *Arabidopsis* (Jones and Dangl, 2006). Next to these intracellular NLR receptors, extracellular LRR classes of R proteins exist (Dangl and Jones, 2001).

The first PTI defence responses of the plant are very fast and occur within minutes, lasting up to several days. These responses are drastic ion-flux changes at the plasma membrane, elevation of cytoplasmic Ca^{2+} levels and ROS production. Ca^{2+} level elevations in the cytosol play a pivotal role in salicylic acid (SA) production, stomatal closure and ROS production. Stomatal closure limits the entry of microbial organisms into leaf tissues. ROS production occurs by activation of the respiratory burst oxidative homolog (RBOH) enzymes in the plasma membrane. These RBOH enzymes are activated directly by RLCKs or by calcium dependent protein kinases (CDPKs). CDPKs are in turn activated by the elevated Ca^{2+} levels and also convey immune signalling to the nucleus resulting in transcriptional reprogramming during PTI. Next to CDPKs, mitogen-activated protein kinase (MAPK) cascades are activated by RLCKs upon MAMP or DAMP perception and transfer the signal to the nucleus to establish PTI by transcriptional reprogramming. Several TFs are regulated by CDPKs and MAPKs and are responsible for this transcriptional reprogramming during PTI, resulting in the production of antimicrobial enzymes or compounds, deposition of callose at the cell wall, cell wall lignification and synthesis of hormones. Callose between the cell wall and the plasma membrane will limit the penetration of microorganisms. Biosynthesis of hormones such as SA, jasmonic acid (JA) and ethylene (ET) is crucial for local and systemic acquired resistances (Muthamilarasan and Prasad, 2013; Couto and Zipfel, 2016).

If the invader produces and secretes effector proteins into the plant cell, PTI can be inhibited by these effectors and the ETI defence responses are triggered. Not all effectors inhibit PTI, some effectors can suppress ETI and some effectors are necessary to change the host metabolism (e.g. to supply nutrients to the pathogen). ETI responses are essentially similar to the responses during PTI, but differ in their strength and duration. Immune responses

during ETI are more prolonged and robust than those during PTI. The ROS production for example is also RBOH dependent, but is of much higher magnitude and sustained upon effector recognition by R proteins (during ETI) (Tsuda and Katagiri, 2010). These similar responses are calcium influx, activation of MAPK cascades, ROS production and transcriptional reprogramming (Dangl and Jones, 2001). ROS are involved in the elimination of the invaders, but also in the activation of transcription. TFs to reprogram transcription are activated by MAPK cascades, next to ROS. They regulate the transcription of several defence genes in and around the infected cell. These defence genes have a function in biosynthesis of hormones (SA, JA and ET), strengthening the cell wall, production of antimicrobial compounds and initiating HR. ETI generates local and systemic acquired resistance (Dangl and Jones, 2001; Tsuda and Katagiri, 2010; Muthamilarasan and Prasad, 2013).

Among the biotic stresses applied on Arabidopsis plants in chapter 3, *Pseudomonas syringae* and *Botrytis cinerea* belong to the phytopathogenic microorganisms or pathogens, *Myzus persicae* belongs to the herbivorous insects. The PTI and ETI response, as well as the hormone signalling, will only be discussed for *P. syringae*, since this is the only pathogen used in the stress tolerance experiments in chapter 4.

1.1.2.2.1. Plant pathogens

Two major groups of plant pathogens can be distinguished. The first group of pathogens first kills the host, usually by the production of toxins, and subsequently feeds on the contents, they are called necrotrophs. In contrast, biotrophs depend on living host tissues to obtain nutrients to complete their life cycle. Pathogens that act as biotrophs or necrotrophs depending on the conditions or the stage of their life cycle, are called hemibiotrophs. Hemibiotrophs act initially as biotrophs, but usually kill the host (necrotrophs) at a later stage of infection (Muthamilarasan and Prasad, 2013; Surico, 2013). Plants, in turn, respond on these pathogens with their sophisticated mechanisms, called the innate immune system (PTI and ETI) resulting in an adaptive response (Dangl and Jones, 2001).

Pseudomonas syringae is a hemibiotrophic bacterium, belonging to the prokaryotic microorganisms (Surico, 2013). *P. syringae* belongs to the gram-negative plant-pathogenic bacteria which can cause several diseases and as such is responsible for major crop yield losses (Büttner, 2016). In crops, seeds infected with *P. syringae* are often the source of disease and disease development is preceded by epiphytic growth of *P. syringae* on the leaf surfaces (Katagiri *et al.*, 2002). Multiple genomes of different strains of *P. syringae* have been sequenced enabling the use of *A. thaliana*-*P. syringae* as a model system for plant-pathogen interactions (Katagiri *et al.*, 2002; Morris *et al.*, 2013). Arabidopsis plants of the ecotype Col are susceptible to *P. syringae* pv. *tomato* DC3000, one of the most widely used virulent strains of *P. syringae*. Disease development starts with the typical water-soaked patches on the leaves which become necrotic and dark-colored later on, surrounded by leaf tissue which shows chlorosis (Katagiri *et al.*, 2002). *P. syringae* can enter the host through physical injuries and natural openings such as stomata (usually on the

leaves) (Katagiri *et al.*, 2002). Once they entered the plant, they inhabit the intercellular spaces and if the plant is susceptible multiply there to high population levels (Tsuda and Katagiri, 2010; Muthamilarasan and Prasad, 2013).

In Figure 1.5 the most important PRRs, regulatory receptor kinases and RLCKs from *A. thaliana* are illustrated and those sensing bacterial PAMPs will be discussed in more detail (Couto and Zipfel, 2016). Chitin is a building block of the fungal cell walls and AtPep1 is a DAMP, an endogenous peptide from *A. thaliana*, derived from PROPEP1, its precursor protein (Yamaguchi *et al.*, 2010; Wirthmueller *et al.*, 2013). Exogenous application of AtPeps (Pep1 - Pep7) has been shown to enhance immunity against *P. syringae* (Yamaguchi *et al.*, 2010).

The bacterial flagellum is important for bacterial pathogenicity in plants (Jones and Dangl, 2006). This flagellum consists of the structural protein bacterial flagellin (Katagiri *et al.*, 2002). Felix *et al.* (1999) showed that a synthetic 22-amino-acid peptide (flg22), a conserved peptide among flagellins of eubacteria, including *P. syringae*, was able to induce many cellular defence responses. Flagellin sensing 2 (FLS2), a PRR from *A. thaliana*, was identified as receptor for bacterial flagellin (Figure 1.5) (Katagiri *et al.*, 2002; Couto and Zipfel, 2016). FLS2 belongs to the LRR-RLKs and upon flg22 perception acts together with a regulatory receptor kinase namely BRI1-associated receptor kinase 1 (BAK1) or somatic embryogenesis receptor kinase 3 (SERK3) to activate several downstream RLCKs such as brassinosteroid-signalling kinase 1 (BSK1), pattern-triggered immunity compromised receptor-like cytoplasmic kinase 1 (PCRK1) and botrytis-induced kinase 1 (BIK1) (Figure 1.5) (Katagiri *et al.*, 2002; Tsuda and Katagiri, 2010; Couto and Zipfel, 2016). Arabidopsis *fls2* mutants revealed enhanced susceptibility to *P. syringae*, indicating the importance of flg22 recognition in plant immunity, limiting the growth of *P. syringae* (Zipfel *et al.*, 2004).

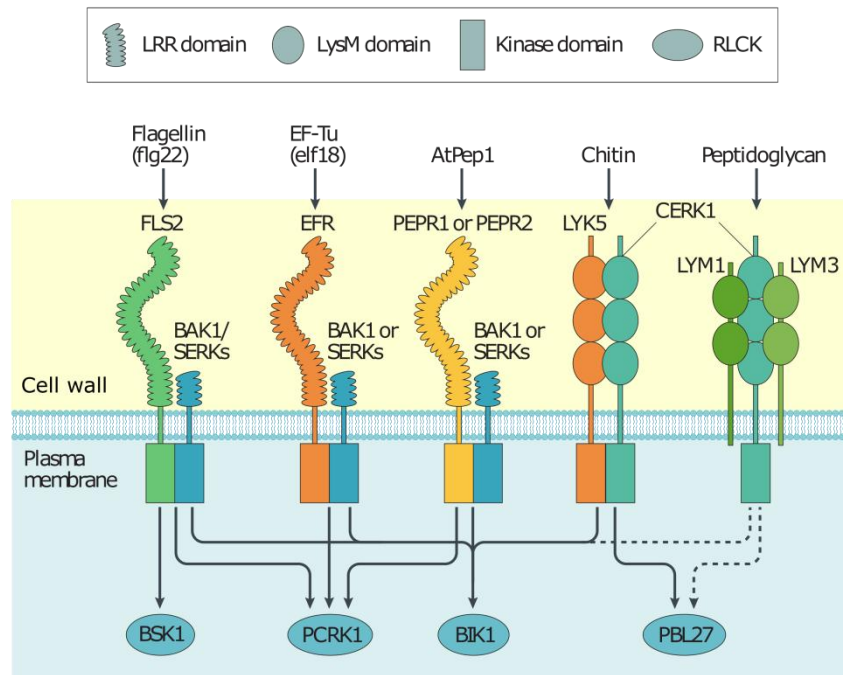


Figure 1.5 Arabidopsis PRRs, regulatory receptor kinases and RLCKs (Adapted from Couto and Zipfel, 2016). Flg22: 22-amino-acid conserved peptide of bacterial flagellin, FLS2: flagellin sensing 2, BAK1: BRI1-associated receptor kinase 1, SERKs: somatic embryogenesis receptor kinases, BSK1: brassinosteroid-signalling kinase 1, PCRK1: pattern-triggered immunity compromised receptor-like cytoplasmic kinase 1, BIK1: botrytis-induced kinase 1, EF-Tu: elongation factor Tu, EFR: EF-Tu receptor, PEPR: Pep receptor, LYK5: LysM-containing receptor-like kinase 5, CERK1: chitin elicitor receptor kinase 1, LYM: LysM domain-containing GPI-anchored protein, PBL27: AvrPphB susceptible 1 (PBS1)-like kinase 27, LRR: leucine-rich repeat, RLCK: receptor-like cytoplasmic kinase.

EF-Tu, the most abundant bacterial protein, is highly conserved in all bacteria, hence also in *P. syringae*. The EF-Tu receptor (EFR) specially recognizes the elf18 peptide which is an N-acetylated peptide consisting of the first 18 amino acids (aa) of the N-terminus of EF-Tu (Kunze *et al.*, 2004; Muthamilarasan and Prasad, 2013). As well as FLS2, EFR interacts with BAK1 or SERK3, the regulatory receptor kinases, upon ligand binding. RLCKs PCRK1 and BIK1 are subsequently activated and take care of the downstream signalling (Figure 1.5) (Couto and Zipfel, 2016). Arabidopsis *efr* mutants revealed a higher efficiency of *Agrobacterium tumefaciens* mediated transferred DNA (T-DNA) transformation, concluding these mutants are more susceptible to *A. tumefaciens* (Zipfel *et al.*, 2006). Kunze *et al.* (2004) compared the alkalization-inducing activity of EF-Tu peptides of *A. tumefaciens* and *P. syringae*, and showed a lower alkalization-inducing activity of the EF-Tu peptide from *P. syringae*.

Peptidoglycan is a major component of the bacterial cell wall (Willmann *et al.*, 2011). LysM domain-containing GPI-anchored proteins (LYMs) from *A. thaliana*, more specific LYM1 and LYM3, contain each two extracellular LysM domains and recruit chitin elicitor receptor kinase 1 (CERK1) during peptidoglycan recognition (Couto and Zipfel, 2016). LYM1 and LYM3 contain no intracellular kinase domain and as such belong to the LysM-RLPs (Muthamilarasan and Prasad, 2013). CERK1 is, equally to BAK1 or SERK3 a regulatory

receptor kinase. Together, LYM1, LYM3 and CERK1 mediate antibacterial immune responses upon binding of peptidoglycan (Figure 1.5) (Couto and Zipfel, 2016). Willmann *et al.* (2011) showed a largely enhanced susceptibility of *lym1*, *lym3* and *lym1 lym3* Arabidopsis mutants to infection with *P. syringae*.

Upon perception of these described PAMPs by their PRRs, PTI is initiated via regulatory receptor kinases and RLCKs (Figure 1.5). Most gram-negative bacteria, like *P. syringae*, deliver their Avr or effector proteins to inhibit PTI to the plant cell by their type III secretion system (T3SS) (Katagiri *et al.*, 2002; Büttner, 2016). This T3SS produces a continuous channel for effectors to be translocated directly into the cytoplasm of the plant cells (Büttner and He, 2009). Next to the interference with PTI, effectors also interfere with signal transduction, proteasome-dependent protein degradation, phytohormone signalling, plant gene expression and the plant cytoskeleton (Büttner, 2016). In total 57 families of effectors were identified in *P. syringae* with each *P. syringae* strain expressing 15 - 30 effector proteins (Lindeberg *et al.*, 2012). Virulent bacteria translocate 15 - 30 effectors via the T3SS into the plants cells. These effectors act as a TF, remodel chromatin and/or affect host TF activity. Generally they inhibit PTI and promote the release of nutrients required for pathogen survival (Feng and Zhou, 2012). For example AvrPphB, a cysteine protease from *P. syringae*, cleaves different Arabidopsis RLCKs, as such inhibiting PTI. One of these RLCKs is PBS1 and the cleavage of PBS1 by AvrPphB is monitored by RPS5, an R protein from *A. thaliana*, initiating ETI. Other RLCKs from *A. thaliana* that can be cleaved by AvrPphB from *P. syringae* are BIK1, PBL1 and PBL2 (Zhang *et al.*, 2010b). Büttner (2016) gives an overview of the targets of several other *P. syringae* Avr or effector proteins.

Hormonal signalling

SA, JA and ET are the classical immune system phytohormones involved in biotic stress responses. SA signalling leads to local and systemic resistance against many biotrophs and hemibiotrophs such as *P. syringae*, while JA and ET signalling promote defence against necrotrophs (Jones and Dangl, 2006; Tsuda and Katagiri, 2010; Couto and Zipfel, 2016). SA and JA mostly work antagonistically (Dangl and Jones, 2001). Note that this is true for *A. thaliana*, but certainly not for all plants. Moreover, other hormones like ABA, gibberellins, auxins, cytokinins and brassinosteroids also play a role in plant immune signalling (Jones and Dangl, 2006; Tsuda and Katagiri, 2010; Pieterse *et al.*, 2012). Only SA perception and signalling will be discussed since this hormone is linked to *P. syringae* infections in Arabidopsis and this is the only biotic stress factor used in the stress experiments in chapter 4.

Biosynthesis of SA is triggered during PTI and ETI upon recognition of PAMPs or effector proteins of biotrophs and hemibiotrophs such as *P. syringae* (Mishina and Zeier, 2007). SA biosynthesis starts from chorismate or phenylalanine via isochorismate synthase (ICS/SID2) or phenylalanine ammonia lyase (PAL), respectively (Garcion and Métraux, 2007). Downstream of SA, expression of several defence-related genes such as

pathogenesis-related (PR) genes is regulated by non-expressor of PR genes 1 (NPR1) (Figure 1.6) (Moore *et al.*, 2011). Several PR genes encode proteins with antimicrobial activity (van Loon *et al.*, 2006). To regulate the expression of these defence-related genes, NPR1 has to be transported to the nucleus via nuclear pore proteins called modifier of *snc1* (MOS) 3, 6 and 7. Only NPR1 monomers can be transported by the MOS nuclear pore proteins (Monaghan *et al.*, 2010).

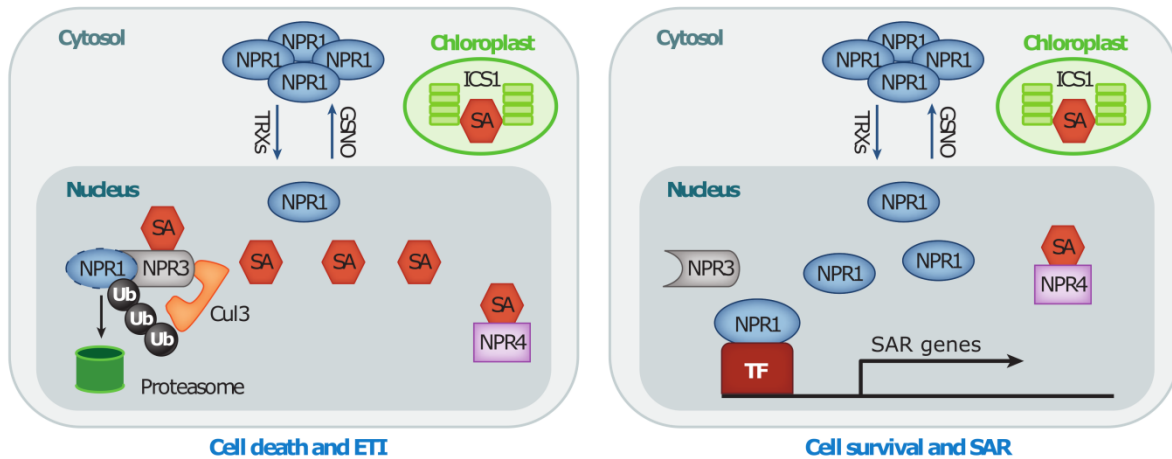


Figure 1.6 Downstream SA signalling (Adapted from Fu and Dong, 2013). Left panel: upon pathogen infection, SA is produced and binds to NPR3, allowing NPR3-mediated degradation of NPR1. This leads to cell death or HR and ETI. Right panel: SA levels in adjacent cells are lower, insufficient to mediate NPR3-NPR1 interaction. NPR4 can still bind SA and as such NPR4-mediated degradation of NPR1 is inhibited. NPR1 monomers accumulate and interact with TFs leading to the activation of different SAR genes. NPR1: non-expressor of PR genes 1, TRXs: thioredoxins, GSNO: S-nitrosoglutathione, SA: salicylic acid, ICS1: isochorismate synthase 1, Cul3: Cullin 3 E3 ligase, NPR3: non-expressor of PR genes 3, NPR4: non-expressor of PR genes 4, ETI: effector-triggered immunity, TF: transcription factor, Ub: ubiquitinylation, SAR: systemic acquired resistance.

In the absence of SA, almost all NPR1 is oligomeric because of intermolecular disulfide bridges and as such cannot be transported into the nucleus (Tada *et al.*, 2008). The NPR1 monomers present can be transported into the nucleus but are ubiquitinated and degraded in the proteasome (Spoel *et al.*, 2009). Changes in the cellular redox state, induced by SA, activate the thioredoxins (TRXs). These TRXs monomerize the oligomeric NPR1, which is subsequently transported into the nucleus (Figure 1.6). S-nitrosoglutathione (GSNO) in contrast facilitates NPR1 oligomer formation in the absence of SA (Tada *et al.*, 2008). NPR3 and NPR4 directly interact with Cullin 3 E3 ligase (Cul3) and as such function as two adaptor proteins mediating the degradation of NPR1. NPR4 is involved in the degradation of NPR1 in the absence of SA, while NPR3 degrades NPR1 upon binding with SA (Figure 1.6). Without pathogen infection, NPR1 is constantly degraded by Cul3 via its interaction with NPR4 (not shown in Figure 1.6). As such unnecessary activation of plant defence is prevented. Upon pathogen infection, SA is produced which can bind to NPR3 and allows NPR3-mediated degradation of NPR1. This leads to HR and ETI (Figure 1.6 left panel). In adjacent cells, the SA levels are lower, too low to mediate the NPR3-NPR1 interaction, but high enough to bind

NPR4 and as such disrupt the NPR4-NPR1 interaction. Neither NPR4 or NPR3 mediate the NPR1 degradation resulting in an accumulation of NPR1 (Fu *et al.*, 2012; Fu and Dong, 2013). Consequently, NPR1 monomers can interact with several members of the TGA subclass of the basic leucine zipper TF family, which in turn can bind promoters of SA-responsive genes (Fan and Dong, 2002). This leads to cell survival and systemic acquired resistance (SAR) (Figure 1.6 right panel) (Fu and Dong, 2013).

Some *P. syringae* strains produce the toxin coronatine which causes tissue chlorosis (Katagiri *et al.*, 2002). Coronatine is a structural mimic of JA and suppresses the SA signalling pathway and thus several defence responses e.g. stomatal closure. The inhibition of the stomatal closure helps *P. syringae* to get access to the apoplast (Jones and Dangl, 2006; Couto and Zipfel, 2016).

1.2 Lectins

The term 'lectin' is derived from 'legere', the Latin word for 'to select' (Van Damme *et al.*, 2008). Indeed, lectins bind selectively to carbohydrate structures. Today, lectins denote all proteins which contain at least one non-catalytic domain that binds reversibly to a specific mono- or oligosaccharide. Lectins are of non-immune origin, found widespread throughout life on earth including bacteria, fungi, viruses, plants and animals (Peumans and Van Damme, 1995; Van Damme *et al.*, 1998).

Plant lectins can be divided in two main classes. One class groups all lectins that are constitutively expressed in high amounts in seeds and vegetative storage tissues. Most of these lectins contain a signal peptide and are as such directed to the secretory pathway. There is evidence that these lectins combine a function as a storage protein with an important role in plant defence against herbivorous insects or animals (Peumans and Van Damme, 1995; Van Damme *et al.*, 1998). The other class contains the nucleocytoplasmic lectins which are expressed in response to certain stress conditions e.g. pathogen attack or environmental changes. In contrast with the abundant vacuolar lectins, these lectins are present in low concentrations in the nucleus and the cytoplasm of the plant cell. Upon stress, the expression of these lectins is elevated, but they are still not that abundant as the vacuolar lectins. Evidence has been presented that these lectins probably interact with glycans at the surface or inside the plant cell and as such play a role in signalling in or between plant cells as part of e.g. plant defence pathways (Lannoo and Van Damme, 2010, 2014).

Most lectins consist of at least one other functional domain, next to their carbohydrate binding domain (Van Damme, 2014). Based on the sequence homology of their carbohydrate binding domain, plant lectins are divided in twelve families. These twelve families are: the *Agaricus bisporus* agglutinin family, the amaranthins, the homologs of class V chitinase-related agglutinin (CRA), the cyanovirin family, the *Euonymus europaeus* lectin

(EUL) family, the *Galanthus nivalis* agglutinin (GNA) family, the hevein family, the jacalin-related lectin (JRL) family, the legume lectin family, the LysM domain family, the *Nicotiana tabacum* agglutinin (Nictaba) family and the ricin B lectin family (Van Damme *et al.*, 2008). Among lectins in the same lectin family, different carbohydrate binding specificities were reported e.g. Nictaba from tobacco and F-box Nictaba from Arabidopsis in the Nictaba family (Stefanowicz *et al.*, 2012). This makes it impossible to classify lectins according to their carbohydrate binding specificity (Van Damme, 2014).

1.3 Nictaba and homologs in *A. thaliana*

1.3.1 Nictaba

Nictaba, the *Nicotiana tabacum* agglutinin, was the first discovered lectin in the Nictaba family (Chen *et al.*, 2002). The Nictaba family groups the nucleocytoplasmic lectins that show sequence homology to the tobacco lectin and is known to be widespread in the plant kingdom (Figure 1.7) (Lannoo and Van Damme, 2010; Delporte *et al.*, 2015; Van Holle *et al.*, 2017a). Nictaba consists out of two identical non-covalently linked subunits of 19 kDa and the determination of its subcellular localization revealed a localization in the nucleus and the cytoplasm of the plant cell (Chen *et al.*, 2002; Lannoo *et al.*, 2006). At plant tissue level, Nictaba is expressed in very young tissues including the apical and root meristems, the cotyledons and the first true leaves (Delporte *et al.*, 2011). Expression of Nictaba was not detectable in tobacco leaves under normal growth conditions. Treatment of the tobacco leaves with methyl jasmonate (MeJA) and insect herbivory revealed a several fold increased expression of Nictaba (Chen *et al.*, 2002; Vandendorre *et al.*, 2009). Lannoo *et al.* (2006) reported the specificity of Nictaba towards *N*-acetyl-D-glucosamine (GlcNAc) oligomers, high-mannose and complex *N*-glycans. Mutational analysis of four conserved amino acid residues in the Nictaba carbohydrate binding site showed that Trp15 and Trp22 play an important role in carbohydrate binding (Schouppe *et al.*, 2010). The identification of core histones as interacting partners for Nictaba and the carbohydrate dependence of this interaction suggest that Nictaba might fulfil a signalling role in response to stress by interacting with *O*-GlcNAcylated histones in the plant cell nucleus (Schouppe *et al.*, 2011; Delporte *et al.*, 2014).

1.3.1 Nictaba homologs in *A. thaliana* (ArathNictabas)

A. thaliana is an important model organism as discussed in section 1.1.1 of this chapter. The genome of *A. thaliana* contains 30 Nictaba related sequences, which are discussed in detail in chapter 2 of this thesis. Nictaba homologs from *A. thaliana* show different domain architectures consisting of a Nictaba domain alone, a combination of a Nictaba domain with an N-terminal Toll/Interleukin-1 receptor (TIR) domain or an N-terminal avirulence induced gene 1 (AIG1)-type G domain or an N-terminal F-box domain (Figure 1.7).

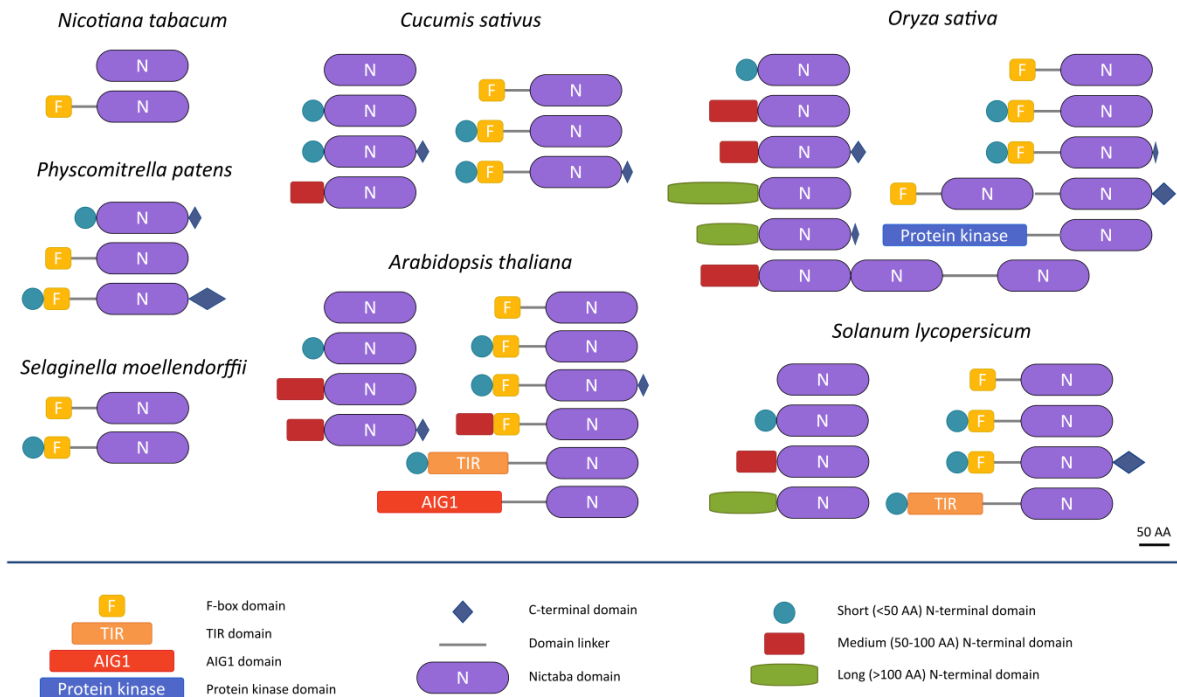


Figure 1.7 Schematic overview of the different domain architectures of Nictaba homologs in different plant species (Updated from Delporte *et al.*, 2015). TIR: Toll/Interleukin-1 receptor, AIG1: avirulence induced gene 1.

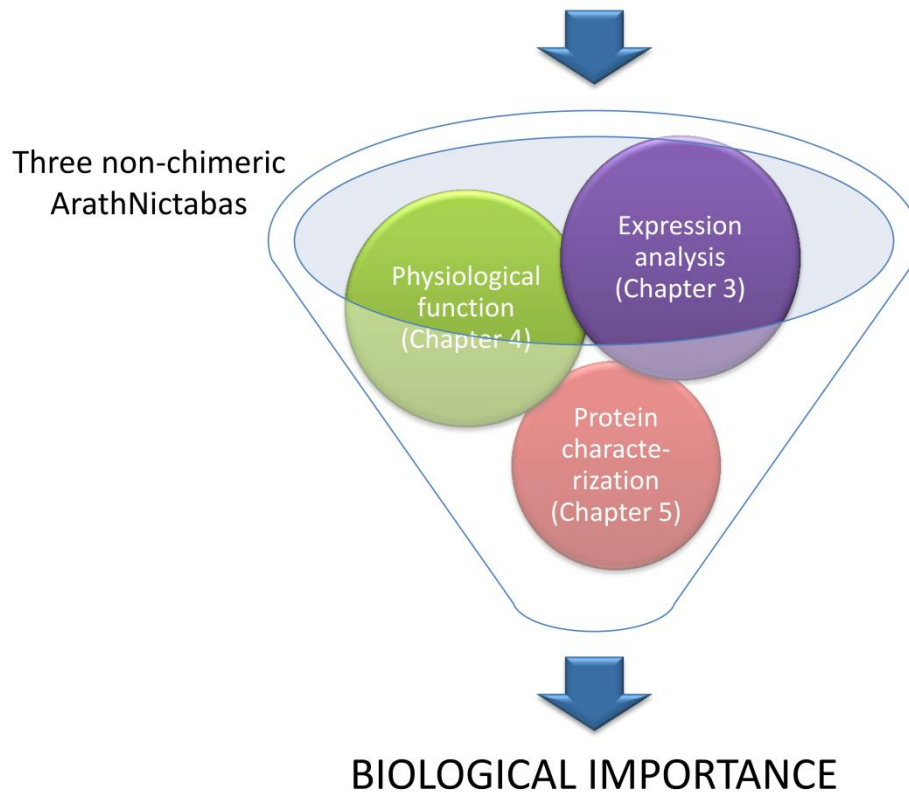
Six Nictaba homologs from *A. thaliana* contain the Nictaba domain alone. Three of these six homologs are the subject of this PhD thesis. Four Nictaba homologs are identified with an N-terminal TIR domain (Eggermont *et al.*, 2017; Chapter 2). TIR domains are frequently found in insects, mammals and plants as part of R proteins of the plant innate immune system. These R proteins contain, next to the TIR domain, a nucleotide binding and LRR domain at the C-terminus and as such are part of the NLR receptors important in initiating ETI (Burch-Smith and Dinesh-Kumar, 2007). Only one Nictaba homolog with an N-terminal AIG1-type G domain was found in *A. thaliana* (Eggermont *et al.*, 2017; Chapter 2). AIG1 domains are also involved in plant defence against pathogens, the name of the domain refers to the *AvrRpt2*-induced gene 1. This R gene is acting to induce ETI upon infection with *Pseudomonas syringae* carrying the *Avr Rpt2* gene, encoding a type III effector protein (Reuber and Ausubel, 1996; Cui *et al.*, 2013). The Nictaba homologs which contain an N-terminal F-box domain are with 19 members the largest group within the Nictaba homolog family of *A. thaliana* (Eggermont *et al.*, 2017; Chapter 2). Proteins containing F-box domains are generally known to play a crucial role in protein degradation using the selective ubiquitin-26S proteasome system (Skaar *et al.*, 2013). More details regarding all these Nictaba homologs and their involvement in plant stress can be found in section 2.4.2.8 of chapter 2 and supplementary table 3.

Scope

Plants are capable to sense changing environmental conditions including drought, extreme temperatures, pathogen infection or insect infestation. In contrast to animals, plants are sessile organisms and as such are not able to escape from these different abiotic and biotic stresses they are confronted with. The only option they have is to fight the adverse environmental conditions by a sophisticated defence system, if successful, leading to adaptation to the stress and survival of the plant. The bad side of this complicated defence system is the energy the plant needs to activate it, energy which cannot be used for growth of the plant. In agriculture plant growth, especially of the vegetative tissues, is important, thus these declines in energy cause tremendous crop yield losses. These huge crop yield losses have to be prevented, certainly with the continuing world population and consumption growth. By 2050, the world population will reach about 9 billion people and recent studies predicted that the world will need 70 to 100 % more food by then. Additionally, climate change will give rise to more plant stress which was not taken into account yet in this prediction. Furthermore in recent decades, due to industrialization, urbanization, desertification and salinization agricultural land that was formerly productive has been lost. At present the land area suitable for agriculture is still decreasing. To cope with this demand of food and to fight additional problems, interest in stress-tolerant crops providing higher crop yields is growing. In order to produce stress-tolerant crops, detailed knowledge of the genetic and biochemical mechanisms underlying plant defence, is necessary. *Arabidopsis thaliana*, the most widely-used model plant, is a valuable system to study these plant defence mechanisms; results can possibly be extrapolated to crops and used to develop modified crops with an improved tolerance to unfavourable environmental conditions.

Plant lectins constitute an important part of this sophisticated defence system, also called the plant innate immune system. Both cell surface localized lectins and intracellular plant lectins play a role in this plant innate immune system. The Nictaba family represents one group of nucleocytoplasmic lectins that contribute to the intracellular signalling that is part of the innate immune system. Nictaba, the *Nicotiana tabacum* agglutinin, was first discovered in 2002, but since then many homologous sequences have been identified in the plant kingdom. Nictaba expression is enhanced after jasmonate treatment and herbivory in the nucleus and the cytoplasm of the plant cell. Inside the nucleus, Nictaba can bind to O-GlcNAc modified histones and is as such believed to remodel the chromatin in order to enhance the transcription of several defence related genes. This PhD work focuses on several Nictaba homologs from *Arabidopsis thaliana* (further referred to as ArathNictabas). The major aim of this PhD thesis is to investigate the biological importance of the ArathNictabas in the stress responses of *A. thaliana* (see schematic overview below).

Screen for lectin genes in *Arabidopsis thaliana* (Chapter 2)



The first objective of this PhD thesis is to screen the *Arabidopsis* genome for the presence of lectin genes from all known plant lectin families, to analyse the protein domains in these lectin sequences and to investigate the phylogenetic relationships (Chapter 2).

The second goal of this PhD work is to unravel part of the physiological function of some non-chimeric ArathNictabas by determining the subcellular localization, expression pattern during plant development and different (a)biotic stress conditions, and tolerance of different ArathNictaba overexpression lines towards different stresses (Chapters 3 and 4).

The third aim of this PhD work is to produce recombinant ArathNictaba protein to determine its lectin activity and carbohydrate specificity, and identify the possible interaction partners for the ArathNictaba proteins (Chapter 5).

Chapter 2

**Genome-wide screening for lectin motifs
in *Arabidopsis thaliana***

This chapter is based on:

Eggermont L, Verstraeten B, Van Damme EJM. 2017. Genome-wide screening for lectin motifs in *Arabidopsis thaliana*. *The Plant Genome* **10**, 1-17.

2.1 Abstract

For more than three decades *Arabidopsis thaliana* served as a model for plant biology research. At present only a few protein families have been studied in detail in Arabidopsis. This study focused on all sequences with lectin motifs in the genome of Arabidopsis.

Based on amino acid sequence similarity, 217 putative lectin genes were retrieved belonging to nine out of twelve different lectin families. The domain organization and genomic distribution for each lectin family was analysed. Domain architecture analysis revealed that most of these lectin gene sequences are linked to other domains, often belonging to protein families with catalytic activity. Many protein domains identified are known to play a role in stress signalling and defence, suggesting a major contribution of the putative lectins in development and plant defence.

This genome wide screen for different lectin motifs will help to unravel the functional characteristics of lectins. In addition, phylogenetic trees and WebLogos were created and showed that most lectin sequences that share the same domain architecture evolved together. Furthermore, the amino acids responsible for carbohydrate binding are largely conserved. Our results provide information about the evolutionary relationships and functional divergence of the lectin motifs in *A. thaliana*.

2.2 Introduction

Proteins are key molecules that fulfil a whole range of biological roles in a cell. Protein domains are distinct parts in the protein sequence that can fold and function separately (Nasir *et al.*, 2014). A protein domain can be defined from different perspectives. The structural viewpoint defines a protein domain as an independent protein fold. From an evolutionary point of view, protein domains are defined as conserved parts of the sequence. Moreover with respect to their function, these domains typically have a particular reoccurring function. All together a protein domain represents a conserved part of the sequence with a specific fold and function (Moore *et al.*, 2008; Kelley and Sternberg, 2015).

The domain architecture of a protein contains all the information of the domains that build the protein and can be determined by scanning the protein sequence through a domain database, such as for example Pfam. Several domains are found at the root of the species tree, indicating that these are common to most species and are used to create a lot of domain architectures by modular rearrangements (Moore *et al.*, 2008). Between 5.6 % and 12.4 % of all currently found domain architectures have been generated more than once throughout evolutionary history. The only reason that this could have happened is because the same domain architectures were formed in different branches of the tree of life as a consequence of selection (Forslund *et al.*, 2008).

Most proteins consist of multiple domains (Rentzsch and Orengo, 2013). Single-domain architectures are common for the major groups of organisms (prokaryotes, eukaryotes and viruses) while multi-domain architectures are usually unique for a species and explain the diversification thereof (Levitt, 2009). Indeed, during the evolution of proteins both domain gains and losses have occurred (Nasir *et al.*, 2014).

Lectins are proteins of which the domain architecture contains at least one lectin domain. A lectin domain can bind reversibly to specific carbohydrate structures either free carbohydrates or glycans from glycoproteins and glycolipids (Peumans and Van Damme, 1995). According to their carbohydrate binding domain, plant lectins can be divided in twelve families: the *Agaricus bisporus* agglutinin family, the amaranthins, the CRA family, the cyanovirin family, the EUL family, the GNA family, the hevein family, the JRL family, the legume lectin family, the LysM domain family, the Nictaba family and the ricin B lectin family (Van Damme *et al.*, 2008). Most lectins contain in addition to their carbohydrate binding domain at least one other functional protein domain (Van Damme, 2014). In principle this protein domain can have a catalytic function. It should be noted that the definition of lectins was established in the early 1980s, and probably needs to be revised in order to take into account the recent developments and novel information with respect to domain organization in chimeric lectins.

For more than three decades *Arabidopsis thaliana* served as a model for plant biology research. At present several reports are accessible on individual proteins containing a lectin domain, but only a few families of proteins have been studied in detail in Arabidopsis. This study aimed to make an inventory of all sequences with lectin motifs in the genome of Arabidopsis. All the lectin sequences from *A. thaliana* were identified and their domain architectures determined. Known functions from literature for several members of each family were discussed. Phylogenetic trees and WebLogos for three lectin families including the JRL family, the LysM domain family and the Nictaba domain family yielded new insights into the phylogenetic relationships of these lectins in plants. All together our study provides information about the evolutionary relationships and functional divergence of the lectin motifs in *A. thaliana*.

2.3 Materials and methods

2.3.1 Identification of the putative lectin genes in *A. thaliana*

Sequences encoding the putative lectin genes were searched for in the *A. thaliana* genome on the Phytozome v10.3 website (<https://phytozome.jgi.doe.gov/pz/portal.html>) (Altschul *et al.*, 1990; Goodstein *et al.*, 2012). Protein sequences encoding model proteins from each lectin family were used as a query for BLASTp searches (Basic Local Alignment Search Tool; comparison matrix BLOSUM62 and word length 3). Each model sequence represents the first lectin sequence described for a particular lectin family, as mentioned in supplementary

table 1. Subsequently, the top hits resulting from these BLASTp searches were again used as a query. All sequences (E-value < 10) were downloaded with the BioMart application from Phytozome v10.3 (Smedley *et al.*, 2015). This high E-value was used to make sure that all putative lectins were retrieved. In case non-lectin domains are selected, they will be deleted in the downstream analysis when the protein domains are annotated (section 2.3.2).

Alternatively, sequences of putative lectin genes in *A. thaliana* were retrieved by using Pfam identification numbers for each lectin domain (Supplementary table 1). However, Pfam identification numbers are not available for the lectin domain of the EUL and CRA family.

Finally, Pfam domain names (Supplementary table 1) were also used to search for potential lectins with the Simple Modular Architecture Research Tool (SMART) database (<http://smart.embl-heidelberg.de/>) (Letunic *et al.*, 2009). Since the SMART database is not using Arabidopsis Gene Initiative (AGI) codes, the AGI codes of the resulting protein sequences were found in the UniProtKB database (<http://www.uniprot.org/uniprot/>).

Identical protein sequences retrieved from the three methods described above were deduplicated.

2.3.2 Annotating the protein domains of putative lectins

All sequences for the putative lectin genes from *A. thaliana* were checked for the presence of a lectin domain with InterProScan5 (<http://www.ebi.ac.uk/interpro>). InterProScan5 scans protein sequences on conserved protein domains and combines data from multiple databases: HAMAP, PANTHER, PfamA, PIRSF, ProDom, PRINTS, Prosite-Profiles, SMART, TIGRFAM, Prosite-Patterns, Gene3d and SUPERFAMILY (Jones *et al.*, 2014). InterProScan5.7-48.0 was installed and ran on the local server. Proteins with at least one lectin domain were considered as a putative lectin. Also protein domains other than lectin domains were identified. The start and end position of each protein domain was determined and used to draw the domain architecture for each putative lectin on scale using the DomainDraw software (<http://domaindraw.imb.uq.edu.au/>) (Fink and Hamilton, 2007).

Since the lectin domains of the EUL and CRA family have no Pfam identification number, the lectin domain sequences for the model sequences (Supplementary table 1) were aligned with the protein sequences encoding the putative lectins from the EUL and CRA family with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers *et al.*, 2011). InterProScan5 was used to identify the domains other than lectin domains in these EUL and CRA family lectins.

2.3.3 Determining signal sequences and/or transmembrane domains

Each potential lectin sequence was checked for the presence of a signal sequence and/or a transmembrane domain with Phobius (<http://phobius.sbc.su.se/index.html>). Phobius combines the models from the SignalP and TMHMM server in a slightly modified way (Käll *et*

al., 2004). The start and end positions for the signal sequences and/or transmembrane domains were determined and used to draw the schematic domain architectures.

2.3.4 Mapping the putative lectin genes on the chromosomes of *Arabidopsis* and analysis of tandem duplications

Using the BioMart application of Phytozome v10.3 the transcription start position (base pairs) of each putative lectin was retrieved and used to map the lectin genes on the chromosomes of *Arabidopsis* (Voorrips, 2002). Only primary transcripts were mapped on the chromosomes. The positions of the centromeres are according to Feraru *et al.* (2012). The chromosomes were drawn to scale using their golden path lengths (TAIR, <https://www.arabidopsis.org/index.jsp>). Tandem duplicated genes were defined as two lectin genes from the same family located on the same chromosome separated by maximum 10 other (not lectin) genes.

2.3.5 Phylogenetic analysis

Phylogenetic trees were created using the lectin domain sequences identified for each family. For putative lectin sequences that contain more than one lectin domain, each lectin domain was used as a separate entry. Protein alignment of these sequences was performed with Multiple Alignment using Fast Fourier Transform (MAFFT, <http://www.ebi.ac.uk/Tools/msa/mafft/>) using the default parameters (Katoh and Standley, 2013). The alignments were trimmed using the automated1 option of trimAl which was installed locally (Capella-Gutiérrez *et al.*, 2009). Unrooted phylogenetic trees were created with RAxML v8.2.4 using the GAMMA model for rate heterogeneity and automatic determination of the best amino acid substitution model (i.e. the model with the highest likelihood score on the starting tree) (Stamatakis, 2014). Bootstrap analysis was performed using the rapid bootstrap algorithm of RAxML (Stamatakis *et al.*, 2008). The number of bootstraps was determined using the frequency criterion up to a maximum number of 1000. Visualization of the phylogenetic tree was done with the FigTree v1.4.3 software (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.3.6 Analysis of amino acids responsible for carbohydrate binding

The untrimmed sequence alignments used to generate the phylogenetic trees were used in WebLogo3 to make a graphical representation of the amino acid conservation at each position of the sequence (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks *et al.*, 2004). A comparative analysis with model sequences and the amino acids known to be important for carbohydrate binding activity allowed to check if amino acids essential for interaction with carbohydrate structures are conserved.

2.4 Results and discussion

2.4.1 Identification and distribution of the genes with a lectin domain in *A. thaliana*

BLASTp searches against the *A. thaliana* genome retrieved 217 putative lectin sequences that could be classified in nine of the twelve plant lectin families as defined in Van Damme *et al.* (2008) (Supplementary table 2). Sequences with lectin domains homologous to the *Agaricus bisporus* agglutinin, amaranthin and cyanovirin families were not found (Table 2.1). The legume lectins represent the most abundant lectin family in *A. thaliana* with 54 putative lectin genes (24.9 %), followed by the JRL family (50 genes; 23.0 %) and the GNA lectin family (49 genes; 22.6 %).

Table 2.1 Predicted lectin sequences and their localization on the chromosomes of *A. thaliana*.

Lectin domain	Putative lectin genes	Percentage	Chromosome location
<i>Agaricus bisporus</i> agglutinin domain	0	0.0	/
Amaranthin domain	0	0.0	/
CRA domain	9	4.1	4
Cyanovirin domain	0	0.0	/
EUL domain	1	0.5	2
GNA domain	49	22.6	1, 2, 3, 4, 5
Hevein domain	10	4.6	1, 2, 3
Jacalin domain	50	23.0	1, 2, 3, 5
Legume lectin domain	54	24.9	1, 2, 3, 4, 5
LysM domain	12	5.5	1, 2, 3, 4, 5
Nictaba domain	30	13.8	1, 2, 3, 4, 5
Ricin B domain	2	1.0	1, 3

Mapping of the transcript start positions on the five chromosomes revealed that putative lectin genes are present throughout the whole *A. thaliana* genome (Table 2.1, Figure 2.1). Genes for some lectin families (e.g. LysM domain family) are present on each chromosome whereas genes from other lectin families (e.g. hevein family) are only present on one or a few chromosomes. The nine putative lectin genes from the CRA lectin family are present in one tandem duplication cluster on chromosome four. Chromosome four is the smallest chromosome and shows also the lowest lectin gene density (1.3 gene/Mbp). Chromosome one is the largest chromosome and has the highest lectin density (2.4 gene/Mbp).

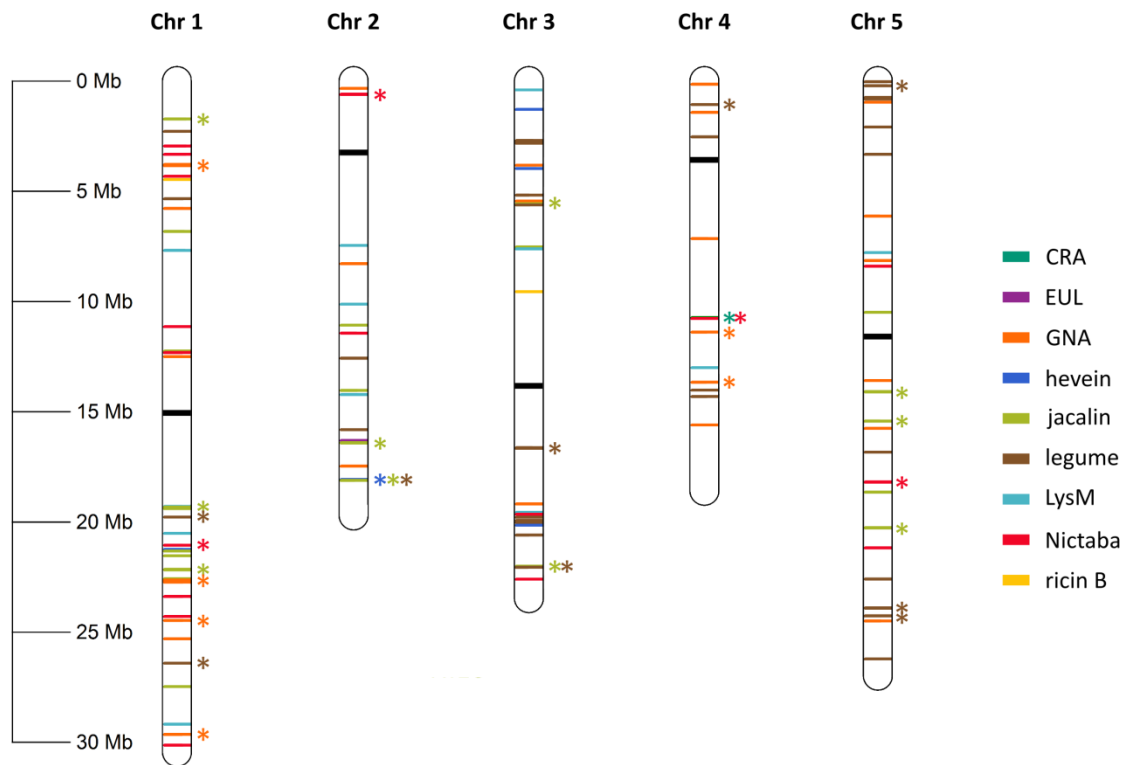


Figure 2.1 Distribution of putative lectin genes on the chromosomes of *A. thaliana*. The chromosomes were drawn to scale. Each lectin family is shown in a specific color. The black bands represent the positions of the centromeres on the chromosomes. Asterisks indicate tandem duplications.

Tandem duplications of lectin sequences are spread throughout the genome and are found for putative lectin genes from six families (Figure 2.1). Chromosome one has the highest amount of tandem duplication clusters (10) whereas chromosome three has the lowest number (4). The legume lectin homologs are the only family with tandem duplication clusters on each chromosome. The GNA family has four tandem duplication clusters on chromosome one, this represents the highest number of tandem duplication clusters on one chromosome.

2.4.2 Domain architecture and importance of the putative lectins in *A. thaliana*

All sequences with a putative lectin domain were also searched for the presence of other protein domains with a known function. Different (combinations of) protein domains can give information on the possible functions of the protein. Different domain architectures for the putative lectins from each family are shown in Figures 2.2, 2.3, 2.4, 2.5, 2.6, 2.7 and 2.8. All sequences were checked for the presence of a signal peptide and/or a transmembrane domain in order to give information about the localization of the putative lectins in the plant cell (Table 2.2). If available, literature reporting on the importance of the lectin sequences for *Arabidopsis* growth and development is briefly discussed. Supplementary table 3 gives an

overview of the publications discussing the biological role(s) of the putative lectins in *A. thaliana*.

Table 2.2 Putative lectin sequences with a signal peptide and transmembrane region.

Lectin family	Signal peptide	Transmembrane domain
CRA	2/9	1/9
EUL	0/1	0/1
GNA	47/49	37/49
Hevein	10/10	0/10
Jacalin	3/50	0/50
Legume lectin	52/54	46/54
LysM	10/12	7/12
Nictaba	0/30	0/30
Ricin B	1/2	0/2

2.4.2.1 CRA homologs

Nine proteins with a CRA domain were found in the *A. thaliana* genome (Table 2.1). All these putative lectin genes are located in one tandem duplication cluster on chromosome four (Figure 2.1). The size of the chitinase-related domain of these sequences varies between 210 and 349 aa with sequence identities to the CRA domain of the model sequence (Supplementary table 1) varying between 41.29 and 55.06 %. Next to their CRA domain, seven out of nine proteins also contain a chitinase insertion domain (CID) (Figure 2.2) (Li and Greene, 2010), a domain typical for chitinases that belong to the glycosyl hydrolase (GH) family 18 (CAZy database). Therefore all nine proteins are classified in the GH family 18 according to this CAZy database. Two out of nine sequences with a CRA domain encode a signal peptide and one of the two proteins contains a transmembrane domain (Table 2.2, Figure 2.2). Transcript levels for one homolog AtChiC (AT4G19810, containing a signal peptide) were slightly upregulated after ABA, JA and salt treatment (Ohnuma *et al.*, 2011). The recombinant AtChiC showed enzymatic activity, in particular it hydrolyzes *N*-acetylglucosamine oligomers (chitinase activity). Taking into account the definition of lectins stating that a lectin domain should not exert any catalytic activity, this protein cannot be referred to as a lectin.

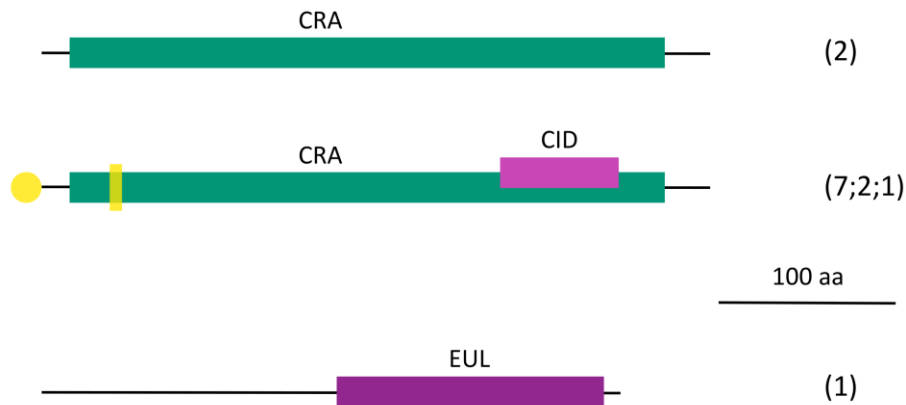


Figure 2.2 Domain architectures of CRA and EUL homologs. Signal peptides and transmembrane regions are drawn in yellow. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively. In case only one number is given, none of the sequences with this domain architecture have a signal peptide or a transmembrane region.

2.4.2.2 *EUL* homologs

Only one EUL homolog was retrieved from the *A. thaliana* genome, referred to as ArathEULS3 (AT2G39050) (Fouquaert *et al.*, 2009). The EUL domain of 154 aa is preceded by an N-terminal domain of 163 aa with unknown function (Figure 2.2). The sequence does not contain a signal peptide nor a transmembrane domain (Table 2.2, Figure 2.2). Microscopic analysis of an enhanced green fluorescent protein (EGFP)-fusion protein revealed that ArathEULS3 is located in the cytoplasm and the nucleus (Van Hove *et al.*, 2011). More recent experiments showed an elevated expression of ArathEULS3 after treatments of Arabidopsis seedlings with glutathione, ABA, MeJA and salt (Hacham *et al.*, 2014; Van Hove *et al.*, 2014). Furthermore Van Hove *et al.* (2015) revealed increased levels of the lectin transcripts after infection of wild type (WT) Arabidopsis plants with *Pseudomonas syringae*. It was suggested that ArathEULS3 plays a role in the ABA-induced stomatal closure (Van Hove *et al.*, 2015).

2.4.2.3 *GNA* homologs

The GNA homologs represent one of the three largest lectin families in *A. thaliana* (Figure 2.3). Only six out of 49 GNA homologs contain only a GNA domain, all other protein sequences have a chimeric domain architecture. The GNA domain is combined with an S-locus glycoprotein domain and/or a Pan/Apple domain and/or a protein kinase domain. In addition, some GNA homologs also possess an S-locus receptor kinase (SRK) domain (Figure 2.3). The domain architecture of one GNA homolog (AT1G11300) contains a tandem repeat of a GNA, S-locus glycoprotein, Pan/Apple and protein kinase domain (not shown in Figure 2.3). Almost all GNA homologs are synthesized with a signal peptide and the majority of them also contain a transmembrane domain (Table 2.2, Figure 2.3).

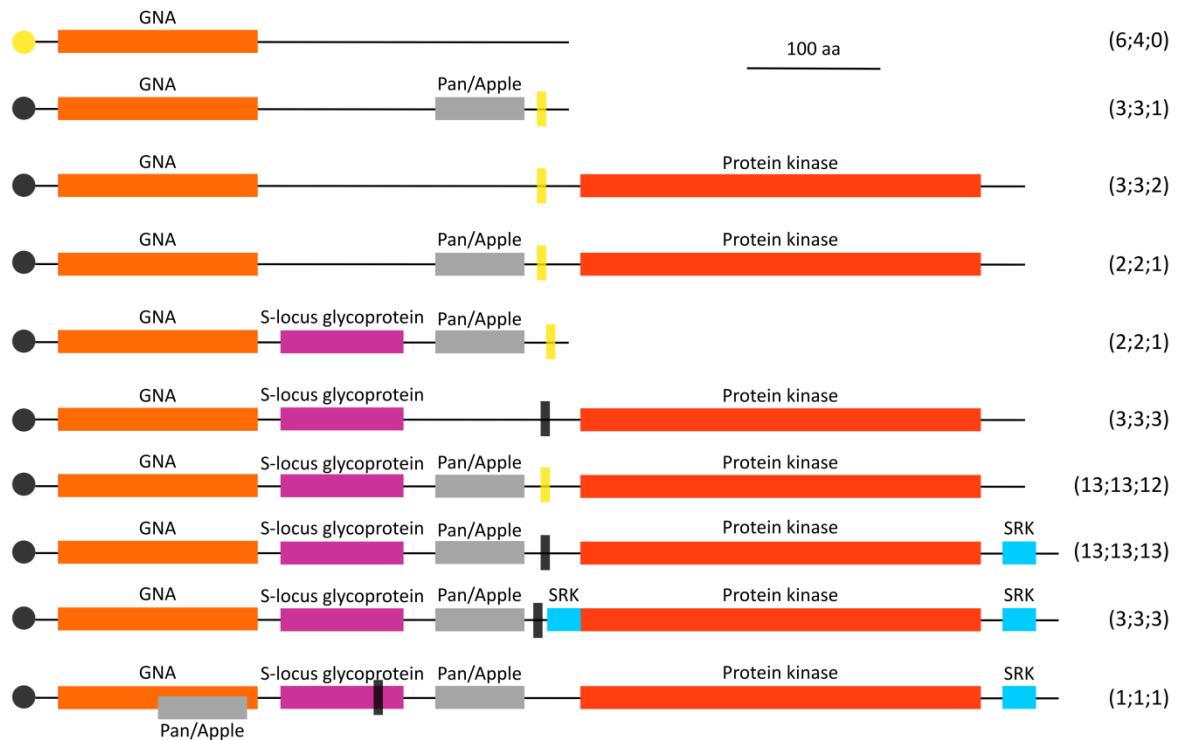


Figure 2.3 Domain architectures of GNA homologs. Signal peptides and transmembrane regions are drawn in black when they are present in all sequences with this domain architecture. They are drawn in yellow when only part of the sequences contains a signal peptide and/or a transmembrane region. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively.

RNA sequencing data suggest the inducible expression for some GNA homologs (AT1G65790, AT5G60900) after inoculation of *Arabidopsis* with the fungal pathogen *Fusarium oxysporum* (Zhu *et al.*, 2013). The expression of some homologs (AT5G60900, AT5G18470), is upregulated in plants exposed to lipopolysaccharides, PAMPs (Sanabria *et al.*, 2008). Lipopolysaccharides are essential components of the bacterial cell wall of gram negative bacteria suggesting that these GNA homologs play a role in the defence against bacterial infections. All but one (AT5G18470) of the GNA homologs that show elevated expression levels after *Fusarium* infection as well as these upregulated by the lipopolysaccharide treatment encode RLKs. Lectin RLKs play a role in plant development, stress and hormonal responses (Vaid *et al.*, 2013). Another RLK (AT4G21390) is more than 200 fold upregulated after ozone (O₃) treatment (Xu *et al.*, 2015). Blaum *et al.* (2014) reported a GNA homolog (AT1G61360) that is co-expressed with BAK1-interacting RLK2, a protein of the LRR-RLKs, which play a role in development and innate immunity. A GNA homolog called calmodulin-binding receptor-like protein kinase 1 (CBRLK1, AT1G11350) harbours a Ca²⁺-dependent CaM binding domain in its C-terminus. This protein possesses autophosphorylation sites, as determined with mass spectrometry (MS) (Kim *et al.*, 2009). All these results suggest that GNA homologs can play a role in different stress related responses.

2.4.2.4 Hevein homologs

Ten hevein homologs were found in the *A. thaliana* genome (Table 2.1). All sequences contain a signal peptide but lack a transmembrane domain (Table 2.2, Figure 2.4) suggesting that all hevein homologs are synthesized following the secretory pathway. All hevein homologs represent chimeric lectin sequences (Figure 2.4), the hevein domain is linked to a chitinase IV domain of the GH family 19 in nine sequences.

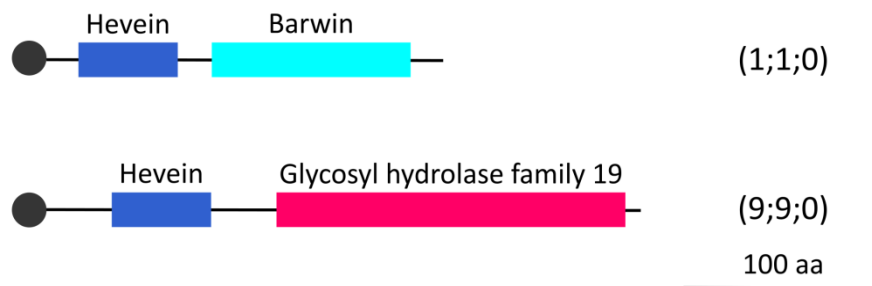


Figure 2.4 Domain architectures of hevein homologs. Signal peptides are drawn in black. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively.

In PR-4 (AT3G04720), the hevein domain is fused to the Barwin domain of 120 aa, named after the Barwin protein of barley which plays a role in the defence against fungal attacks (Ludvigsen and Poulsen, 1992). Genes that encode the PR proteins are rapidly induced after pathogen attacks and treatment with certain hormones. Recent studies revealed that the expression of PR genes is also regulated by environmental factors like light and abiotic stresses. For example, transcript levels for PR-3 (AT3G12500), a hevein homolog without Barwin domain, and PR-4 are significantly higher after high salt treatment and respond in an ABA-dependent manner (Seo *et al.*, 2008). These genes are also responsive to sulfur dioxide exposure and are involved in the defence mechanism against the fungal pathogen *Alternaria brassicicola* (Thomma *et al.*, 1999; Mukherjee *et al.*, 2010; Li and Yi, 2012). Price *et al.* (2015) studied the hevein domain of some class IV chitinases from Arabidopsis by MALDI-TOF MS analysis and revealed the presence of three conserved disulfide bridges in the hevein domain as reported for the model lectin hevein. A whole genome microarray revealed that the expression of another hevein homolog (AT3G54420), called AtEP3, is upregulated after treatment of Arabidopsis with NO (Parani *et al.*, 2004). AtEP3 plays a role in programmed cell death (Passarinho *et al.*, 2001).

2.4.2.5 *Jacalin homologs*

The jacalin homologs represent a large group of putative lectins in *A. thaliana*. In contrast to many other lectin families in Arabidopsis, most of the jacalin homologs are composed of one or more jacalin domains only (Figure 2.5). Only 6 of the 50 jacalin homologs are chimeric lectins. Four of them contain two to four Kelch motives, a 44–56 amino acid motif that first was discovered in *Drosophila* and forms a single four-stranded antiparallel β -sheet. Proteins containing kelch repeats play a role in many aspects of cell function (Adams *et al.*, 2000). The remaining two chimeric jacalin homologs contain an F-box associated domain (type 1) and/or an F-box domain. The F-box motif consists of approximately 60 aa and links the F-box protein to the SCF complex involved in protein degradation (Kipreos and Pagano, 2000). Only three of the JRLs contain a signal peptide and none of them has a transmembrane domain, suggesting that most of the jacalin homologs are cytoplasmic proteins (Table 2.2, Figure 2.5).

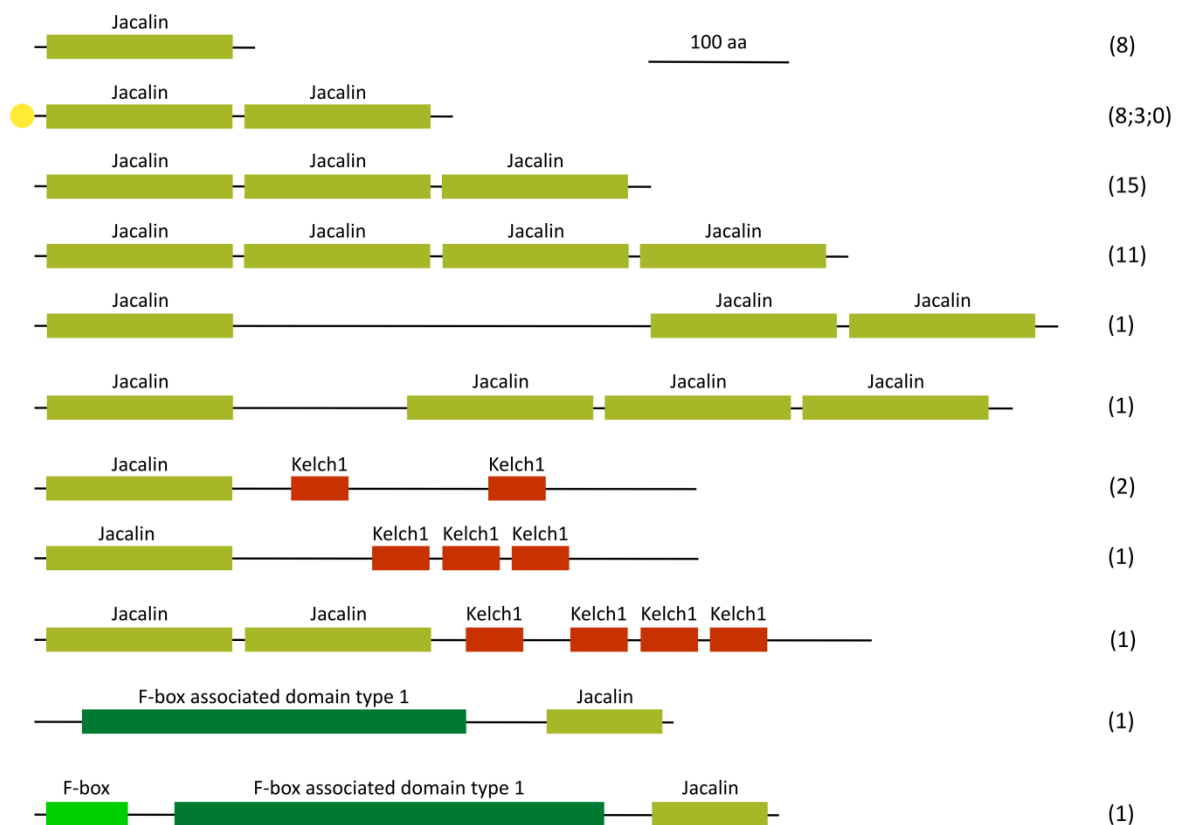


Figure 2.5 Domain architectures of jacalin homologs. Signal peptides are drawn in yellow. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively. In case only one number is given, none of the sequences with this domain architecture have a signal peptide or a transmembrane region.

Since *pbp1* (PYK10 binding protein 1, AT3G16420) mutants show yellow phenotypes de Luna-Valdez *et al.* (2014) suggested that the jacalin homolog PBP1 is involved in the development of chloroplasts. PBP1 may also act as a molecular chaperone that helps the

correct polymerization of PYK10 in response to tissue damage and destruction of subcellular structures. PYK10 is a β -glucosidase of the myrosinase family (Nagano *et al.*, 2005). β -glucosidases hydrolyse glucosinolates, a group of secondary plant metabolites that play a key role in the myrosinase-glucosinolates plant defence system. This hydrolysis only takes place upon tissue damage as the myrosinases and glucosinolates have a different subcellular localization. The product of the hydrolysis is an aglucone that can spontaneously rearrange into an isothiocyanate. Isothiocyanates are particularly toxic for microorganisms, nematodes and insects (Nagano *et al.*, 2005; Wittstock and Burow, 2010). Two jacalin homologs are myrosinase binding proteins (MBPs) referred to as MBP1 (AT1G52040) and MBP2 (AT1G52030) (Capella *et al.*, 2001). These MBPs interact with myrosinases in order to form large complexes (Nagano *et al.*, 2005). Four other jacalin homologs are known as *A. thaliana* nitril specifier proteins (AtNSP1-4, respectively AT3G16400, AT2G33070, AT3G16390, AT3G16410). They catalyze the formation of nitriles from the aglucone (Kong *et al.*, 2012). The jacalin homolog AtJAC1 (AT3G16470) plays a role in controlling the flowering time of *A. thaliana*. AtJAC1 regulates the expression of the repressor gene flowering locus C (FLC) through interaction with glycine-rich RNA-binding protein 7 (GRP7) and as such influences the flowering time (Xiao *et al.*, 2015). The jacalin homolog restricted tobacco etch potyvirus (TEV) movement 1 (RTM1, AT1G05760) restricts the long-distance movement of TEV in the phloem by acting together with the non lectin proteins RTM2 and RTM3. As such the plant can prevent systemic infection (Chisholm *et al.*, 2001). The RTM system can also act towards the plum pox virus (PPV) and the lettuce mosaic virus (Revers *et al.*, 2003; Decroocq *et al.*, 2006). The expression of the jacalin homolog encoded by AT1G52000 is upregulated after the inoculation of *Arabidopsis* with the fungal pathogen *Fusarium oxysporum* (Zhu *et al.*, 2013).

2.4.2.6 Legume lectin homologs

The legume lectin homologs represent the largest group of putative lectins in *A. thaliana* (Table 2.1). Although 54 homologs are found, their domain architecture is relatively simple (Figure 2.6). Only 13 of the 54 homologs contain the legume lectin domain alone, 41 legume lectin homologs contain a protein kinase domain C-terminally linked to the legume lectin domain. Except for two legume lectin homologs all putative legume lectins contain a signal peptide, 46 of them also contain a transmembrane domain (Table 2.2). All homologs that contain a protein kinase domain are synthesized with a signal peptide and contain a transmembrane domain.

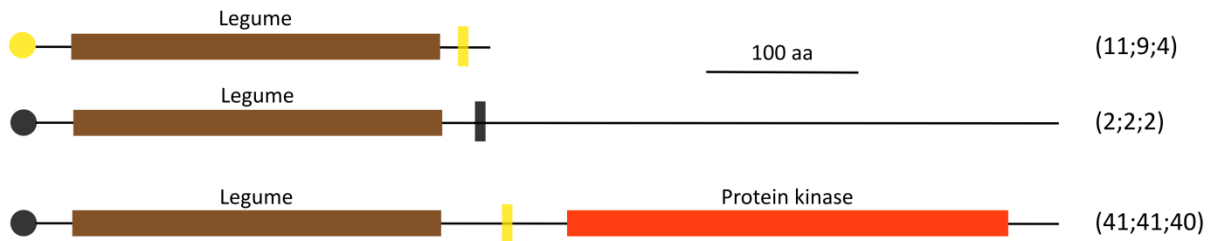


Figure 2.6 Domain architectures of legume lectin homologs. Signal peptides and transmembrane regions are drawn in black when they are present in all sequences with this domain architecture. They are drawn in yellow when only part of the sequences has a signal peptide and/or a transmembrane region. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively.

Mutation of the SGC Lectin RLK (AT3G53810), a legume lectin homolog with a protein kinase domain, resulted in *Arabidopsis* plants with male sterility as a result of a defect in pollen development (Wan *et al.*, 2008). Several legume lectin homologs (all with a protein kinase domain) are stress related proteins. The expression of the legume lectin homolog LecRK-b2 (AT1G70130), for example, was upregulated by ABA, salt and osmotic stress (Deng *et al.*, 2009). He *et al.* (2004) demonstrated that the expression of AtLecRK2 (AT3G45410) is elevated upon salt stress and this response is regulated by the ET signalling pathway. The expression of legume lectin homolog AtLPK1 (AT4G02410) is highly upregulated after ABA, MeJA and SA treatments. Overexpression of this legume lectin homolog in *A. thaliana* showed a better seed germination under high salt conditions (Huang *et al.*, 2013). The legume lectin homolog named LecRK-VI.2 or LecRKA4.1 (AT5G01540) plays a role in the ABA stress response and the disease resistance of *Arabidopsis* against *Pseudomonas syringae* and *Pectobacterium carotovorum* (Xin *et al.*, 2009; Singh *et al.*, 2012). Also legume lectin homologs, LecRKA4.2 (AT5G01550) and LecRKA4.3 (AT5G01560) have a role in the ABA stress response. These legume lectin homologs are located next to each other on chromosome five (Xin *et al.*, 2009). The expression of some of these legume lectin homologs is also influenced by *A. brassicicola*, *F. oxysporum* and ozone (Mukherjee *et al.*, 2010; Zhu *et al.*, 2013; Xu *et al.*, 2015). LecRK-I.9 (AT5G60300) plays a role in the cell wall - plasma membrane adhesions. The destabilization of these adhesions is one of the ways used by *Phytophthora brassicae* to infect *Arabidopsis*. Overexpression of LecRK-I.9 in *A. thaliana* resulted in enhanced resistance to *P. brassicae* suggesting that LecRK-I.9 contributes to strengthening of cell wall - plasma membrane adhesions (Bouwmeester *et al.*, 2011).

2.4.2.7 *LysM* homologs

The *LysM* homologs represent a small group in *A. thaliana* with diverse domain architectures (Figure 2.7). In five out of twelve sequences the *LysM* domain is linked to a protein kinase domain. One of these sequences contains two *LysM* domains. According to InterProscan one homolog with a protein kinase domain possesses a syndecan/neurexin domain. Although

this homolog is well described in literature, the syndecan/neurexin domain has never been reported. One of the twelve LysM homologs contains an N-terminal F-box domain. Half of the sequences encoding LysM homologs consist of one or two LysM domains. No other protein domains were identified in these coding sequences. Ten LysM homologs are synthesized with a signal peptide and seven of them contain a transmembrane domain, indicating that most of the LysM homologs follow the secretory pathway (Table 2.2).

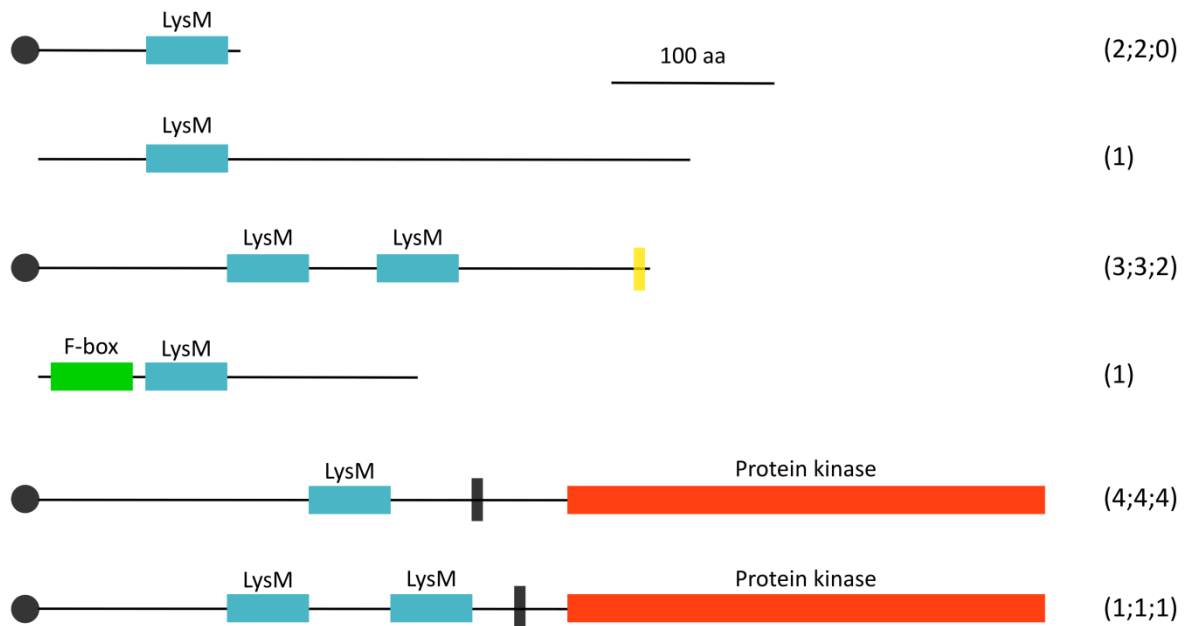


Figure 2.7 Domain architectures of LysM homologs. Signal peptides and transmembrane regions are drawn in black when they are present in all sequences with this domain architecture. They are drawn in yellow when only part of the sequences has a signal peptide and/or a transmembrane region. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively. In case only one number is given, none of the sequences with this domain architecture have a signal peptide or a transmembrane region.

Zhang *et al.* (2007) described five of the LysM homologs that contain a protein kinase domain. One of these homologs, referred to as AtCERK1, AtLYK1 or LysM RLK1 (AT3G21630) is important in the response of Arabidopsis to fungi. Mutation of RLK1 created more susceptible Arabidopsis plants to fungal pathogens (Wan *et al.*, 2008). More recent research revealed that the expression of AtLYK4 (AT2G23770) and AtLYK5 (AT2G33580) is upregulated by chitin. Pull-down analysis also proved that these LysM homologs can interact with chitin. Only mutants of AtLYK4 showed a reduced expression of chitin-responsive genes (like *atlyk1* mutants) and a higher susceptibility to *Alternaria brassicicola* and *Pseudomonas syringae*. Since AtLYK4 does not possess an active protein kinase domain, it is hypothesized that AtLYK1 and AtLYK4 may form a chitin-receptor complex with a single active kinase domain (AtLYK1) that starts the downstream chitin signalling (Wan *et al.*, 2012). Furthermore,

AtLYK3 (AT1G51940) may be important for the cross talk between the ABA and pathogen stress response (Paparella *et al.*, 2014).

2.4.2.8 *Nictaba* homologs

Among the 30 *Nictaba* homologs retrieved from the *Arabidopsis* genome, 19 sequences contain an F-box domain (Figure 2.8). Four *Nictaba* homologs contain an N-terminal TIR domain, a domain with a role in pathogen detection and defence responses (Burch-Smith and Dinesh-Kumar, 2007). Only one putative *Nictaba* lectin contains an AIG1-type G domain, a domain which is also found in GTPases that play a role in the defence against pathogens (Reuber and Ausubel, 1996). The absence of signal peptides or transmembrane domains suggests a cytoplasmic localization for the *Nictaba* homologs.

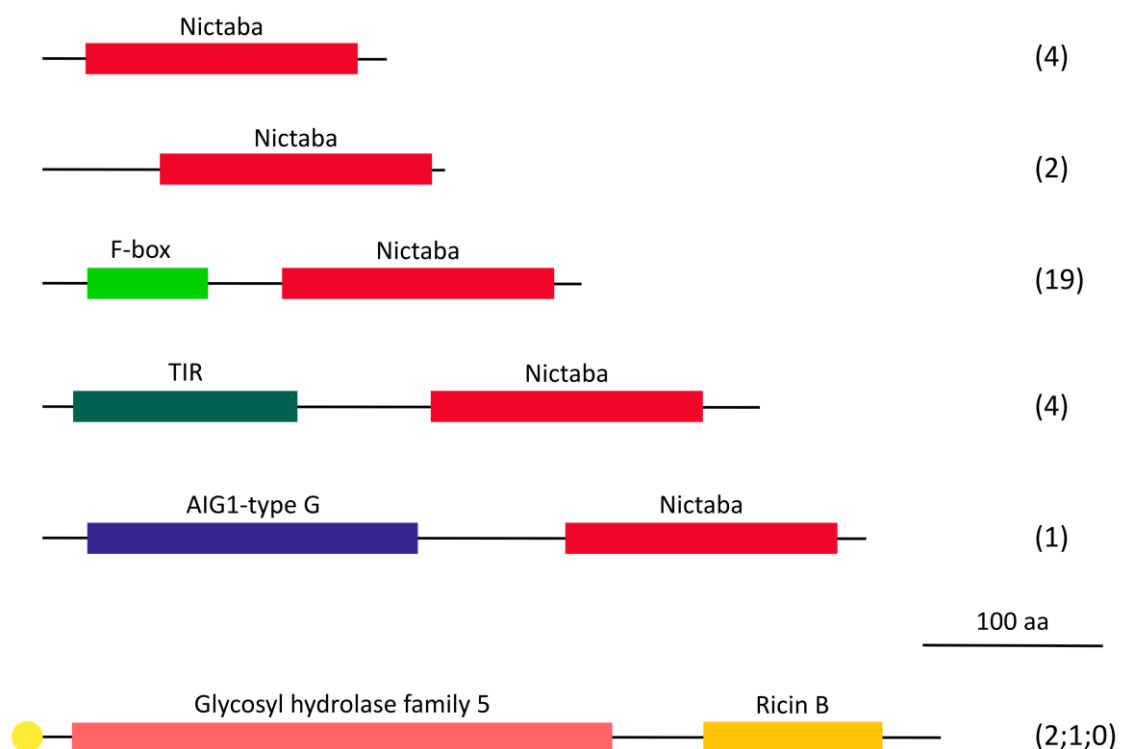


Figure 2.8 Domain architectures of *Nictaba* and ricin B homologs. Signal peptides are drawn in yellow. The numbers in brackets indicate the total number of sequences with this domain architecture, followed by the number of sequences that contain a signal peptide and a transmembrane region, respectively. In case only one number is given, none of the sequences with this domain architecture have a signal peptide or a transmembrane region.

Phloem protein 2 (PP2)-A1 (AT4G19840), a *Nictaba* homolog that contains only a *Nictaba* domain, is part of the phloem protein bodies in the sieve elements. Recombinant protein production and glycan array analysis demonstrated binding of PP2-A1 to *N*-acetylglucosamine oligomers, high-mannose *N*-glycans and 9-acyl-*N*-acetylneuraminic sialic acid. Although PP2-A1 did not show insecticidal properties against *Acyrtosiphon pisum* and *Myzus persicae*, the weight gain of the nymphs was reduced by adding the recombinant

PP2-A1 protein to the aphid diet (Beneteau *et al.*, 2010). According to Lee *et al.* (2014) PP2-A1 showed molecular chaperone as well as antifungal activity. Since the expression of PP2-A1 is upregulated after pathogen attack and ET treatment, the molecular chaperone can play a crucial role in the stress response (Lee *et al.*, 2014). More than half of the F-box containing Nictaba homologs were shown to interact with at least one Arabidopsis Skp1-like protein (Gagne *et al.*, 2002; Risseuw *et al.*, 2003; Takahashi *et al.*, 2004; Dezfulian *et al.*, 2012; Kuroda *et al.*, 2012). A lot of F-box containing Nictaba homologs are proven to be stress-inducible, mostly by abiotic stresses but to a lesser extent also by biotic stresses (Takahashi *et al.*, 2004; Dezfulian *et al.*, 2012; Kuroda *et al.*, 2012). Transcript levels for the VIP1 binding F-box protein (VBF, AT1G56250) are upregulated by *Agrobacterium tumefaciens* and help to bring the T-DNA inside the plant by degrading VirE2 and VIP1 proteins that coat the T-DNA (Zaltsman *et al.*, 2010). The expression of PP2-B11 (AT1G80110) is elevated by salt stress and overexpression lines of this gene are more tolerant to high salinity conditions (Jia *et al.*, 2015). Lee *et al.* (2014) showed that overexpression lines of PP2-B11 are more sensitive to drought stress suggesting that PP2-B11 is a negative regulator in the response to drought stress. Carbohydrate binding activity was reported for the recombinant PP2-B10 or F-box Nictaba (AT2G02360) protein expressed in *Pichia pastoris* (Stefanowicz *et al.*, 2012). Glycan array analysis demonstrated that F-box Nictaba recognizes *N*-acetylglucosamine, Lewis A, Lewis X, Lewis Y and blood type B motifs. It is hypothesized that F-box Nictaba plays a role in nucleocytoplasmic protein degradation (Lannoo *et al.*, 2008). Recently Stefanowicz *et al.* (2016) showed that overexpression of this F-box Nictaba resulted in a reduction of leaf damage upon infection with *Pseudomonas syringae*. Another Nictaba homolog (AT3G61060) with an F-box domain is downregulated during callus initiation (Xu *et al.*, 2012).

2.4.2.9 Ricin B homologs

Only two ricin B homologs exist in *A. thaliana* (Table 2.1). They both have the same domain architecture consisting of an N-terminal GH family 5 domain and a C-terminal ricin B domain (Figure 2.8). Only one homolog is synthesized with a signal peptide and none of the sequences has the characteristics of a transmembrane domain (Table 2.2). The expression of one of the ricin B homologs (AT3G26140) is upregulated after infection of the plants with PPV (Babu *et al.*, 2008).

2.4.3 Phylogenetic analysis and analysis of the carbohydrate binding site.

Screening of the Arabidopsis genome for lectin sequences and analysis of the protein domains yielded a lot of different protein domain combinations composed of a lectin domain linked to the AIG1-type G, Barwin, CID, F-box, F-box associated domain type 1, GH family 5, GH family 19, Kelch1, Pan/Apple, protein kinase, S-locus glycoprotein, SRK or TIR domain. Strikingly the protein kinase and the F-box domains are found in combination with different lectin domains. All the putative lectins with a protein kinase domain are synthesized with a signal peptide and almost all of them have a predicted transmembrane region, suggesting

that these chimeric proteins will follow the secretory pathway and will probably reside in the plasma membrane or in the apoplast. All putative lectin sequences with an F-box domain do not possess a signal peptide nor a transmembrane region, and presumably encode nucleocytoplasmic proteins. Although the role of the F-box domain is known, research on genes containing this F-box domain fused to a lectin domain (either a JRL, a LysM or a Nictaba domain) is far from being fully implemented. No information is available with respect to the physiological importance of the F-box JRLs and the F-box LysM, but several F-box Nictaba proteins were shown to be stress related proteins and as such may play a role in plant defence. To check for evolutionary relationships between F-box-lectin domain combinations, phylogenetic trees were created for each of these lectin families. In case a sequence contains multiple lectin domains they were separated. For instance the first jacalin domain of AT1G19715.1 is designated AT1G19715.1_1. In addition, the carbohydrate binding site of all the lectin domains was analysed to check the conservation of the amino acids responsible for interaction with the carbohydrate.

2.4.3.1 Evolutionary relationships of the jacalin homologs from *A. thaliana*

The family of JRLs represents a large group of putative lectins in *A. thaliana*. Most sequences consist of one or more jacalin domains, only 12 % of the sequences encode chimeric proteins. The dendrogram constructed with the individual jacalin domain sequences from all putative JRLs yielded a very complex tree. As shown in Figure 2.9 both sequences containing the F-box associated domain type 1 (AT3G59590.1 and AT3G59610.1) are very closely related. All JRL domains associated with two or more Kelch1 domains are grouped together except for the N-terminal jacalin domain of the only Kelch1 JRL sequence containing two jacalin domains (AT3G16410.1_1). The bootstrap value of the branch is quite high (93 %) suggesting that the jacalin domain sequences that occur in combination with Kelch1 domains evolved together whereas the first jacalin domain of AT3G16410.1 is more similar to the jacalin domains of the non-chimeric JRLs.

Most JRLs composed of a single jacalin domain are separated in two small clusters (Figure 2.9). The majority of the lectin sequences contain multiple jacalin domains. Overall, two big clades of jacalin domains belonging to sequences with multiple JRL domains can be distinguished (Figure 2.9). Clade A (blue) contains all the first jacalin domains from most non-chimeric JRLs and all second jacalin domains except from one sequence belonging to the JRLs containing four jacalin domains. Clade B (purple) contains all the other jacalin domains of the non-chimeric JRLs and the jacalin domains of the F-box associated domain type 1 and the Kelch1 containing sequences. Interestingly jacalin domains from JRLs containing multiple lectin domains usually do not cluster together in the tree, except for the three jacalin domains of AT1G19715.1. These data suggest that tandemly arrayed jacalin domains within one JRL are not formed by a duplication event. Three JRLs composed of two tandem jacalin domains are synthesized with a signal peptide. The first and second jacalin domain of these three sequences also group together (Figure 2.9). Multiple small tandem duplication clusters

are found throughout the tree, grouping closely related jacalin domain sequences (Figure 2.1, Figure 2.9).

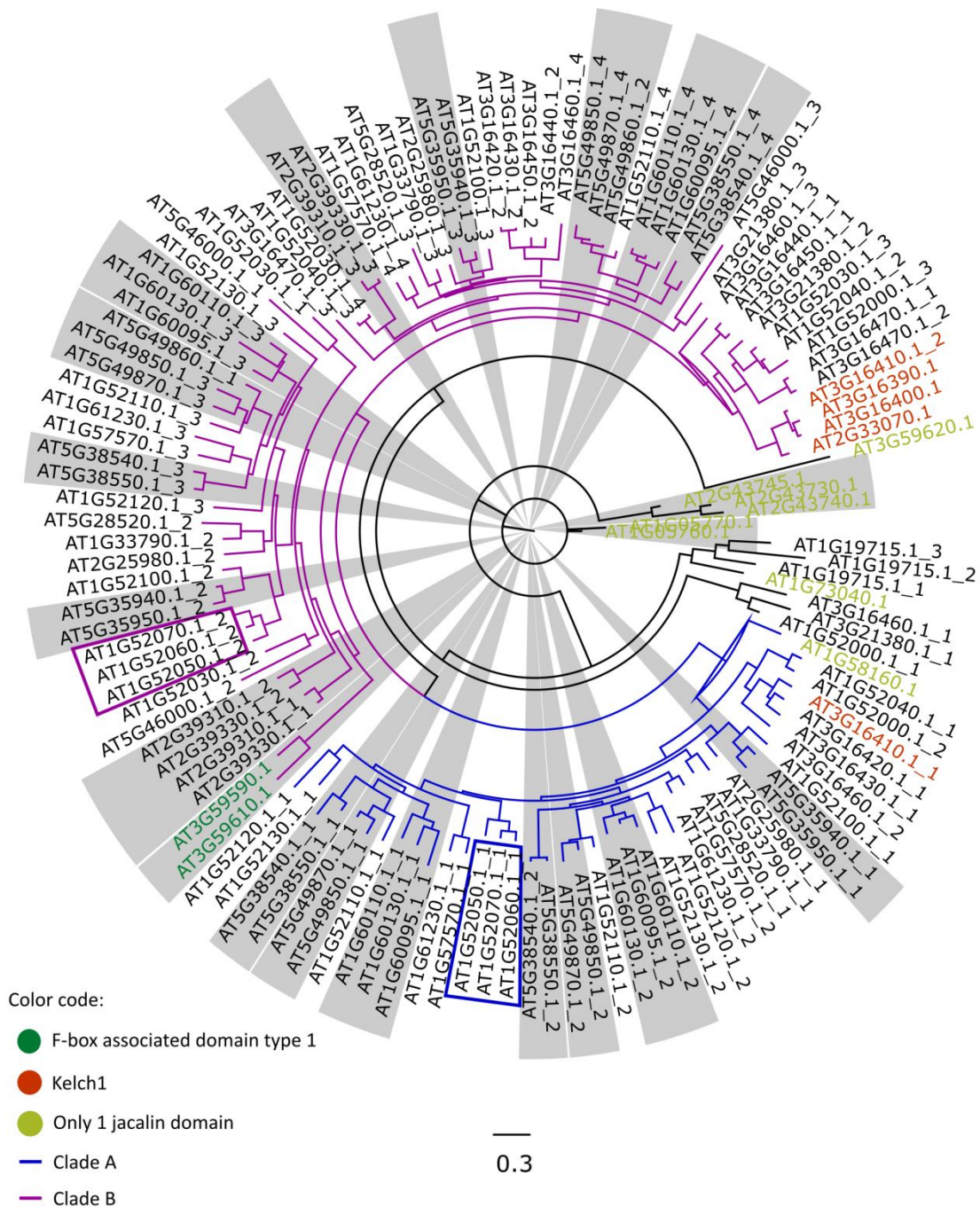


Figure 2.9 Maximum likelihood phylogenetic tree of jacalin domains. The blue and the purple frame mark the first as well as the second jacalin domains of the three sequences with a signal peptide, respectively. Small tandem duplication clusters are highlighted in grey. The coloured branches represent clade A (blue) and clade B (purple). The scale bar represents the mean of the number of substitutions per site according to a maximum likelihood estimation.

2.4.3.2 Conserved amino acids in the carbohydrate binding site of jacalin homologs from *Arabidopsis*

The first JRL, referred to as jacalin, was identified from the seeds of jackfruit *Artocarpus integrifolia* (Bunn-Moreno and Campos-Neto, 1981). Jacalin is composed of four subunits, two α -chains (133 aa) and two β -chains (20 aa), and exhibits specificity towards galactose (Sankaranarayanan *et al.*, 1996; Houless Astoul *et al.*, 2002). Since the discovery of jacalin, many jacalin homologs with specificity towards galactose and mannose have been identified throughout the plant kingdom. In contrast to the galactose specific jacalins that are confined to the Moraceae, the mannose specific jacalins are widespread in higher plants (Houless Astoul *et al.*, 2002). Next to the difference in carbohydrate binding specificity, galactose and mannose specific jacalins differ in the maturation of the lectin polypeptide and their localization in the cell. The post-translational cleavage of the lectin precursor into the α - and β -chain does not take place in the mannose specific jacalins. Whereas galactose specific JRLs reside in the vacuole, mannose specific JRLs are cytoplasmic proteins (Peumans *et al.*, 2000). Houless Astoul *et al.* (2002) re-investigated the carbohydrate binding specificity of jacalin and concluded that the specificity of jacalin is not restricted to galactose and *N*-acetylgalactosamine, but extends to mannose and glucose. They reported that the carbohydrate binding site of jacalin consists of an N-terminal glycine and three C-terminal amino acids namely Tyr, Trp and Asp. However, only the N-terminal glycine and the C-terminal aspartic acid were found in the WebLogo resulting from the protein alignment of 127 jacalin domain sequences from *Arabidopsis* (Figure 2.10).

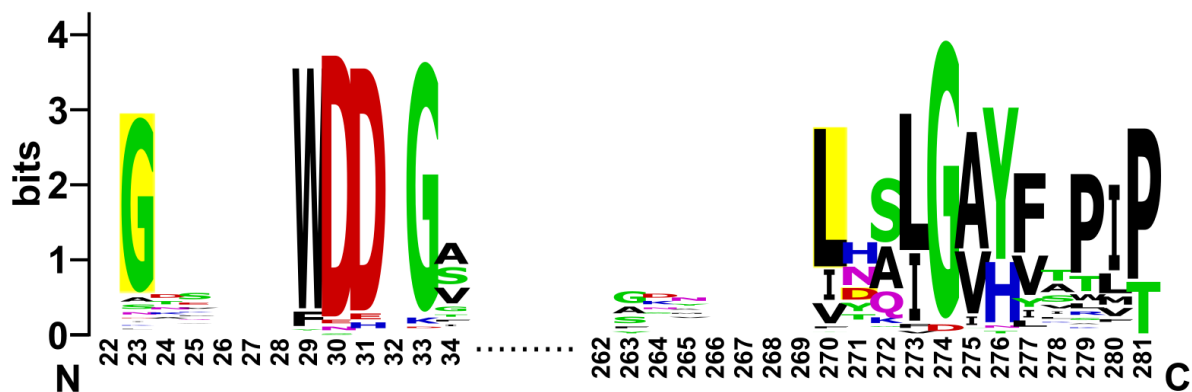


Figure 2.10 WebLogo of amino acids responsible for carbohydrate binding in the JRL domains. Each column of the alignment is represented by a stack of letters. The height of each letter indicates the frequency of this letter at that position of the sequence. The height of the overall stack is proportional to the sequence conservation. The conserved amino acids important for carbohydrate binding activity are highlighted in yellow. Note that the leucine residue is largely conserved, but was not shown to be important for carbohydrate binding activity. Only the regions around the conserved amino acids are shown. The complete WebLogo can be found in supplementary data (Supplementary figure 1).

Judging from the height of the residue, it can be concluded that the N-terminal glycine is conserved for most of the sequences encoding jacalin homologs from *Arabidopsis* whereas

the C-terminal aspartic acid is not well conserved. The complete WebLogo shows a lot of gaps in the alignment, making it difficult to interpret the data (Supplementary figure 1).

Bourne *et al.*, (2004) reported the residues important for carbohydrate binding in the *Calystegia sepium* agglutinin, a mannose specific JRL. The carbohydrate binding site consists of an N-terminal glycine (Gly17), Asn96, Tyr141, Tyr142 and Asp144. Also these residues were not conserved in the JRLs from Arabidopsis. The jacalin domains from Arabidopsis have a well conserved leucine residue preceding the C-terminal aspartic acid needed for carbohydrate binding (Figure 2.10). Although this residue was not shown to be important for carbohydrate binding Raval *et al.*, (2004) also noticed that this leucine is largely conserved in 58 sequences of individual jacalin domains from JRLs belonging to 16 different plant species.

2.4.3.3 *Phylogenetic analysis of the LysM domain family in A. thaliana*

The LysM family represents a small group of putative lectins in Arabidopsis. In addition to the LysM domain, only two protein domains have been identified, namely an F-box domain and a protein kinase domain. In the phylogenetic tree constructed with the sequences of the LysM domains only, the LysM domain that is associated with the F-box domain forms a separate group with the LysM domain from AT5G23130.1 (Figure 2.11), representing the only LysM sequence that is synthesized without a signal peptide (similar to F-box-LysM sequence). This LysM domain is linked to a C-terminal sequence with unknown function. The LysM domains that are associated with the protein kinase domains are found in different branches of the tree. The N-terminal LysM domains of the three sequences with a tandem array of LysM motifs cluster together in one branch of the tree (Figure 2.11).

2.4.3.4 Conservation of the carbohydrate binding site in the LysM homologs from *Arabidopsis*

LysM domains are very short lectin motifs of approximately 40 aa. The first LysM domain was discovered in the lysozyme from *Bacillus* phage ϕ 29 (Garvey *et al.*, 1986). In plants, most LysM proteins for which carbohydrate binding activity is known, belong to the group of LysM receptor kinases. These LysM modules can bind peptidoglycans from bacterial pathogens and chitin fragments from fungal pathogens (Lannoo and Van Damme, 2014). Plant LysM modules recognize chitin fragments (GlcNAc)_n when the degree of polymerization is higher than five ($n \geq 5$) (Petutschnig *et al.*, 2010). Kitaoku *et al.* (2016) investigated the (GlcNAc)_n binding site of a LysM domain containing protein from the green algae *Volvox carteri* composed of two N-terminal LysM domains (96 % sequence homology) and a C-terminal catalytic domain. The binding of chitin fragments to the second LysM domain involved a hydrophobic interaction between Trp96 of the LysM domain and the pyranose ring of (GlcNAc)_n. NMR-based titration experiments revealed that other amino acid residues were also important to form the carbohydrate binding site: Gly92, Asp93, Thr94, Phe95, Ala97, Ile98, Ala99, Gln100, Ala119, Arg120, Leu121, Gln122 and Gly124 (Kitaoku *et al.*, 2016). The corresponding residues Asp9, Thr10, Ile14 and Ala15 were conserved in the LysM domains from *A. thaliana* (Figure 2.12). Malkov *et al.* (2016) described the chitin fragment binding site of the LySM homolog LYR3 from legumes as a hydrophobic tunnel consisting of Pro, Phe, Tyr, Leu and Trp residues. Similarly, the hydrophobic amino acids Tyr3, Leu11, Leu47 and Pro50 are highly conserved in the LysM domains from *A. thaliana*. Buist *et al.* (2008) created a WebLogo using all existing LysM containing proteins from the Pfam database. It can be concluded that the conserved residues reported by these authors are highly similar to the conserved amino acids from the LysM motifs in *A. thaliana*, as shown in the WebLogo resulting from the protein alignment of 16 LysM domain sequences (Figure 2.12).

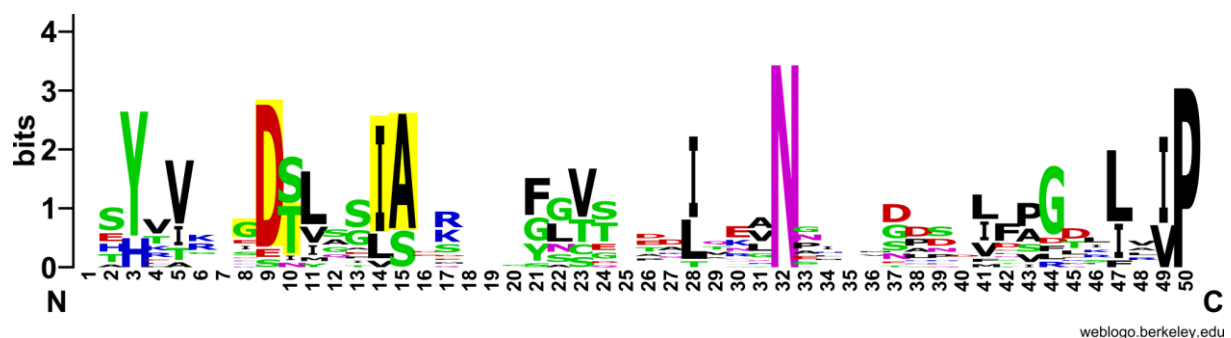


Figure 2.12 WebLogo of amino acids responsible for carbohydrate binding in the LysM domains. Each column of the alignment is represented by a stack of letters. The height of each letter indicates the frequency of this letter at that position of the sequence. The height of the overall stack is proportional to the sequence conservation. The conserved amino acids important for carbohydrate binding activity are highlighted in yellow.

2.4.3.5 Phylogenetic analysis of the Nictaba family in *A. thaliana*

The Nictaba family groups 13.8 % of all putative lectins from *A. thaliana*. Most Nictaba related sequences contain an N-terminal F-box domain linked to a Nictaba domain, a domain architecture that is also abundant in several crop species (Delporte *et al.*, 2015; Van Holle *et al.*, 2017a). The phylogenetic tree built from the Nictaba sequences can be divided in two big clades (A and B) (Figure 2.13). Clade B mainly contains F-box Nictaba sequences. Only AT2g02280.1 does not possess an F-box domain, but has an N-terminal region consisting of less than 10 aa preceding the Nictaba domain. Maybe At2g02280.1 lost its F-box domain during evolution. Clade A clusters all Nictaba sequences that occur in combination with the TIR domain or the AIG1-type G domain, the Nictaba sequences composed only of a Nictaba domain and a small group of F-box Nictaba homologs. This clustering suggests that the Nictaba domains of the non-chimeric sequences are more closely related to the TIR and AIG1-type G Nictaba sequences. Three out of four TIR Nictaba sequences occur in a tandem duplication on chromosome five and group together in the dendrogram (Figure 2.13). The subgroup of F-box Nictaba sequences in clade A differs from most F-box Nictaba sequences in clade B in that the latter sequences belong to two tandem duplication clusters on chromosomes 1 and 2 (Figure 2.13).

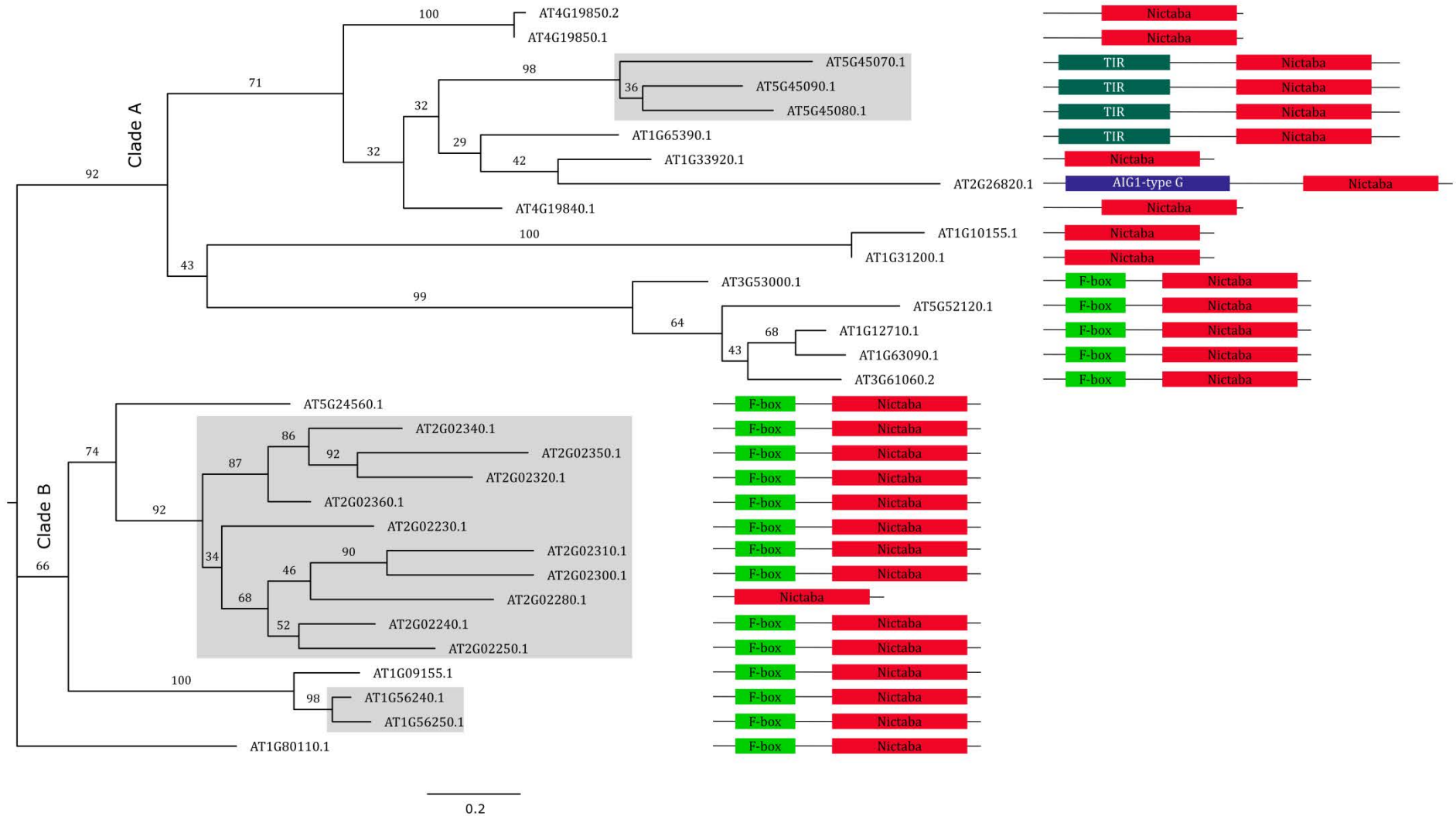


Figure 2.13 Maximum likelihood phylogenetic tree of Nictaba domains. The domain architecture of every putative lectin is represented using the same color code as in Figure 2.8. Tandem duplication clusters are highlighted in grey. The bootstrap values are given at the corresponding branches. The scale bar represents the mean of the number of substitutions per site according to a maximum likelihood estimation.

2.4.3.6 Conserved amino acids in the carbohydrate binding site of Nictaba homologs from *Arabidopsis*

Three-dimensional modelling of the Nictaba sequence from *Nicotiana tabacum* suggested that the amino acids Trp15, Trp22, Glu138 and Glu145 of Nictaba are important for interaction with the carbohydrate. Mutational analysis indicated that only the tryptophan residues play an important role in the carbohydrate binding site of Nictaba (Schoupe *et al.*, 2010). The WebLogo made for 31 Nictaba domain sequences from *Arabidopsis* revealed that the four residues identified for the lectin from tobacco are quite conserved (Figure 2.14). However, as can be observed in the complete WebLogo (Supplementary figure 2) there are a lot of gaps in the protein alignment. It should also be noted that the conserved residues in the Nictaba sequences from *Arabidopsis* are similar to the ones reported for the Nictaba domains from soybean (Van Holle *et al.*, 2017a). The fact that several amino acids in the binding site are conserved does not allow to draw conclusions with respect to the carbohydrate-binding activity of these proteins. Stefanowicz *et al.* (2012) reported that the specificity of F-box Nictaba (At2G02360) and Nictaba are different, despite the fact that the amino acids responsible for carbohydrate binding are conserved. These results suggest that other amino acids in the vicinity of the binding site also play a role in the conformation of the binding site and its interaction with glycans.

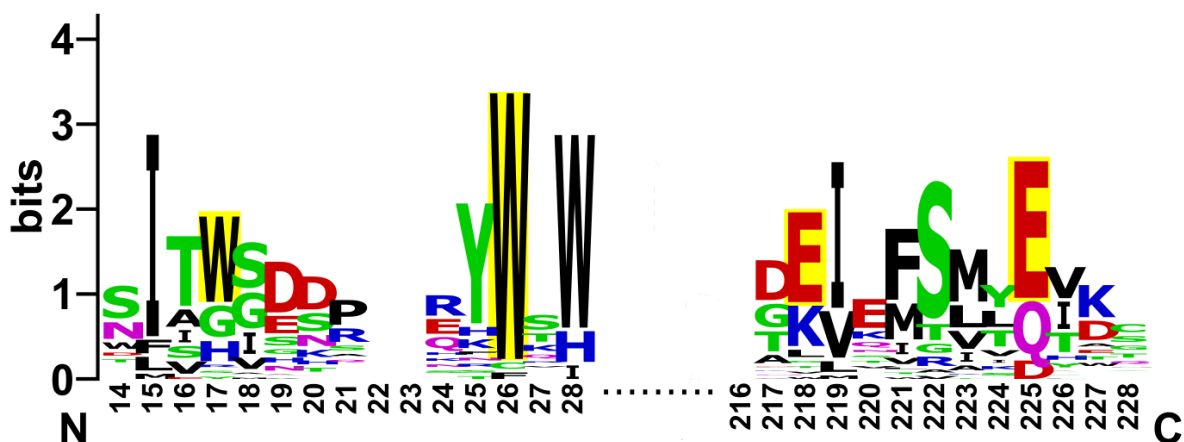


Figure 2.14 WebLogo of amino acids responsible for carbohydrate binding in the Nictaba domains. Each column of the alignment is represented by a stack of letters. The height of each letter indicates the frequency of this letter at that position of the sequence. The height of the overall stack is proportional to the sequence conservation. The conserved amino acids important for carbohydrate binding activity are highlighted in yellow. Only the regions around the conserved amino acids are shown. The complete WebLogo can be found in supplementary data (Supplementary figure 2).

2.5 Conclusions

A total of 217 putative lectin sequences were identified in the genome of *A. thaliana*. Analysis of the domain architectures for each sequence revealed that most sequences contain multiple protein domains. Most of the known protein domains associated with lectin motifs have been reported to be involved in stress signalling, development and defence. Although some domain architectures are unique for a specific lectin domain (AIG1-type G, Barwin, CID, F-box associated domain type 1, GH family 5, GH family 19, Kelch1, Pan/Apple, S-locus glycoprotein, SRK and TIR domain), we also retrieved some protein domains that are associated with multiple lectin motifs, in particular the F-box and protein kinase domain.

Judging from the absence or the presence of signal peptides and/or transmembrane regions in the lectin sequences it is obvious that the putative lectins will end up in different locations in the cell. Taken into account the ambiguity in lectin specificity for many carbohydrate binding recognition domains, our data do not allow drawing conclusions with respect to the activity of the carbohydrate binding site. However, literature data confirm that at least some lectin motifs exert carbohydrate binding activity and are involved in protein-carbohydrate interactions.

Taken together all information it is clear that *A. thaliana* plants have at their disposal a whole range of proteins with lectin motifs. These putative lectins are located in different locations in the cell and in different plant tissues, and can exert complementary activities. It can be envisaged that these lectins play an important role in plant development and survival under stress conditions.

Chapter 3

**Comparative analysis of the Nictaba
homologs in *Arabidopsis thaliana***

3.1 Abstract

Nictaba is a nucleocytoplasmic lectin originally identified in tobacco leaves. Expression of Nictaba is upregulated after jasmonate treatment and insect herbivory. Within the genome of *A. thaliana*, six non-chimeric Nictaba homologs were identified, consisting of an N-terminal sequence and a Nictaba domain. In this chapter, a thorough expression analysis is presented of three non-chimeric Nictaba homologs from *A. thaliana* to elucidate whether these Nictaba homologs have biological properties similar to the lectin from tobacco. This includes the subcellular localization and the expression of the Nictaba homologs in different tissues during plant development. Furthermore, the stress responsiveness of three genes encoding non-chimeric Nictaba homologs was investigated. Two Nictaba homologs show a nucleocytoplasmic localization while the third homolog is only localized to the cytoplasm. qRT-PCR analysis revealed expression of all three Nictaba homologs in different tissues throughout the development. All three Nictaba homologs show a different expression pattern upon different stresses. Taken together, our data suggest that the Nictaba homologs represent stress regulated proteins and possibly play a role in plant stress responses.

3.2 Introduction

Plant lectins can be divided in two main classes. The first class groups the vacuolar lectins, these proteins are present in high amounts in seeds and vegetative storage tissues. Next to their function in nitrogen storage, these vacuolar lectins have an important role in plant defence (Van Damme *et al.*, 1998). The second class contains the nucleocytoplasmic lectins which are usually present in low concentrations in normal growth conditions but their expression is induced upon plant stress. Evidence has been presented that these lectins play a role in the signal transduction as part of plant defence pathways (Lannoo and Van Damme, 2010, 2014).

Six out of the twelve plant lectin families (Chapter 2) contain inducible, nucleocytoplasmic lectins (Lannoo and Van Damme, 2010). One of these six families is the Nictaba family which is known to be widespread in the plant kingdom (Delporte *et al.*, 2015; Van Holle *et al.*, 2017a). Nictaba, the *Nicotiana tabacum* agglutinin, was the first discovered lectin in this family (Chen *et al.*, 2002). Part of the research in our group focussed on Nictaba from *N. tabacum*, an F-box Nictaba homolog from *A. thaliana* and some Nictaba homologs from *Glycine max* and revealed more information about the role of Nictaba homologs in the plant stress response.

Nictaba, the tobacco lectin, consists of two identical non-covalently linked subunits of 19 kDa (Chen *et al.*, 2002). Glycan array screening revealed the specificity of Nictaba towards GlcNAc oligomers, high-mannose and complex *N*-glycans (Lannoo *et al.*, 2006). Based on molecular modelling and protein sequence alignment Schouppe *et al.* (2010) deduced that Trp15, Trp22, Glu138 and Glu145 are conserved amino acid residues in the Nictaba binding

site, suggesting that these residues play a role in sugar binding. Mutational analysis of these four residues showed that the tryptophan residues play an important role in the carbohydrate binding site of Nictaba (Schouppe *et al.*, 2010). Nictaba expression was not detectable under normal conditions but increased several fold after plant exposure to stress situations such as insect herbivory and jasmonate treatment (Chen *et al.*, 2002; Vandenborre *et al.*, 2009). Nictaba is located in the nucleus and the cytoplasm of the plant cell (Chen *et al.*, 2002). Delporte (2013) and Lannoo *et al.* (2006) verified the nucleocytoplasmic localization of Nictaba using EGFP fusion constructs transiently and stably transformed in different plant systems. Using *A. thaliana* plants stably expressing a Nictaba promoter- β -glucuronidase fusion construct, Delporte *et al.* (2011) showed promoter activity in very young tissues including the apical and root meristems, the cotyledons and the first true leaves. Expression of Nictaba in the roots after MeJA and cold stress treatment of tobacco was confirmed using ELISA and Western blot analysis (Delporte *et al.*, 2011). The identification of core histones as interacting partners for Nictaba and the carbohydrate dependence of this interaction suggest that Nictaba might fulfil a signalling role in response to stress by interacting with *O*-GlcNAcylated histones in the plant cell nucleus (Schouppe *et al.*, 2011; Delporte *et al.*, 2014).

The genome of *A. thaliana* contains 30 Nictaba homolog sequences, identified in chapter 2. A large number of these Nictaba homologs (19/30) have an N-terminal F-box domain, next to their carbohydrate binding domain (Eggermont *et al.*, 2017; Chapter 2). In this study our aim is to focus on the non-chimeric Nictaba homologs from *A. thaliana* and to investigate whether these Nictaba homologs have similar biological properties as the lectin from tobacco.

A first step to elucidate the biological relevance of these *A. thaliana* Nictaba homologs is a thorough expression analysis. This includes the localization in the plant cell and the expression of the Nictaba homologs in different tissues during plant development. Furthermore, the stress responsiveness of the expression of these non-chimeric Nictaba homologs was investigated.

3.3 Materials and methods

3.3.1 Plant material and growth conditions

WT *A. thaliana* seeds, ecotype Col-0, were purchased from Lehle Seeds (Round Rock, Texas, USA). Arabidopsis seeds were grown in pots containing commercial soil or individually grown in artificial soil (Jiffy-7, 44 mm \emptyset , distributed by InterGrow, Aalter) in a controlled growth chamber at 21 °C with a 16/8 h light/dark photoperiod after a 3 days stratification period at 4 °C in the dark. The light intensity in the controlled growth chamber was approximately 100 $\mu\text{mol}/\text{m}^2\cdot\text{s}$ (Radium Spectralux plus white (58W) lamps). Alternatively, seeds were grown *in vitro*, therefore seeds were surface sterilized in 70 % ethanol for two minutes followed by

5 % bleach for 10 minutes. Afterwards, the seeds were rinsed four to eight times with sterile distilled water until the pH of the water was approximately neutral. The sterilized *Arabidopsis* seeds were sown *in vitro* on solid Murashige and Skoog (MS) medium (4.3 g/L MS salts with vitamins and nutrients (Duchefa), 30 g/L sucrose (Applichem), pH 5.7-5.8 (adjusted with 0.5 M NaOH) and 8 g/L plant agar (Duchefa)) with a selective antibiotic if necessary. After a 3 days stratification period at 4 °C in the dark to eliminate any residual dormancy of the seeds, the plates were transferred to a growth chamber at 21 °C with a 16/8 h light/dark photoperiod.

WT *N. benthamiana* seeds were supplied by dr. Verne A. Sisson (Oxford Tobacco Research Station, Oxford, NC, USA). For transient transformation, the tobacco seeds were sown in pots containing commercial soil and cultivated in a controlled growth chamber at 25 °C with a 16/8 h light/dark photoperiod. All plants in soil were watered regularly.

3.3.2 *In silico* tools

Multiple protein sequence alignments were performed with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and pairwise protein sequence alignments were obtained using EMBOSS Water (http://www.ebi.ac.uk/Tools/psa/emboss_water/).

Prediction of the presence of a signal peptide was conducted using Phobius and SignalP 4.1 (Käll *et al.*, 2004; Petersen *et al.*, 2011) whereas the presence of a nuclear localization signal (NLS) was predicted by NucPred (Brameier *et al.*, 2007). Using SUBA3, protein subcellular localization was predicted (Hooper *et al.*, 2014). Nucleotide BLAST (BLASTn) and translated nucleotide BLAST (tBLASTn) searches against the expressed sequence tag (EST) database were performed using the NCBI website (<https://blast.ncbi.nlm.nih.gov>).

Promoter sequences were downloaded from the Phytozome database (<https://phytozome.jgi.doe.gov>) and cis-acting regulatory promoter elements were predicted using PlantCARE (Lescot *et al.*, 2002) and AGRIS (Yilmaz *et al.*, 2011). Using the eFP browser (Winter *et al.*, 2007) and Genevestigator (Hruz *et al.*, 2008), microarray expression data were analyzed to study the developmental expression as well as the expression under different stress conditions. Microarray data from the eFP browser representing transcript levels under different stress conditions (Winter *et al.*, 2007) were visualized in a heat map using the BAR HeatMapper Plus Tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi).

3.3.3 Expression in normal growth conditions during plant development

For the gene expression analysis experiments under normal growth conditions, different tissues were sampled during plant development.

For the aerial plant tissues, WT *Arabidopsis* seeds were sown *in vitro* on MS medium (until 22 days) or in Jiffy's. The *Arabidopsis* seeds were stratified and transferred to a growth

chamber as described in section 3.3.1. The *Arabidopsis* plants grown in Jiffy's were watered regularly and fertilizer was added once after 25 days. Whole plantlets were collected at 6, 15 and 22 days after sowing the seeds. Rosette leaves from at least five plants were harvested and pooled after 31 days. After 39 and 54 days, rosette leaves, cauline leaves, stems and flowers from at least five plants were sampled and pooled.

For the root samples, WT *Arabidopsis* seeds were sown in expanded clay granules ($\varnothing < 4$ mm). The *Arabidopsis* seeds were stratified and transferred to a growth chamber as described in section 3.3.1. The *Arabidopsis* plants were watered regularly and fertilizer was added once per week. Root samples from at least 20 plants were collected and pooled after 34, 46 and 59 days.

All samples were immediately frozen in liquid nitrogen and stored at -80 °C prior to RNA extraction and quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) analysis. Two biological replicates were performed and analyzed, each with two technical replicates.

3.3.4 Hormone and abiotic stress treatments

Sixteen-day-old *Arabidopsis* seedlings grown *in vitro* on filter paper on top of MS medium were used to treat with the following solutions: 100 μ M MeJA, 100 μ M ABA, 300 μ M SA and 150 mM NaCl. For each treatment, the filter papers with the germinated seedlings were transferred to Petri dishes filled with liquid MS medium containing either the hormone or the salt solution and incubated at 21 °C. Prior to use, stock solutions of the hormones (MeJA, ABA and SA) were made in 100 % ethanol and water was used in case of the salt stock solution. Control plants were kept on liquid MS medium containing an equal concentration of the corresponding solvent (ethanol or water). Heat stress was applied by incubating the plates with seedlings in the dark at 37 °C, controls were incubated at 21 °C in the dark. For every treatment, fifty seedlings were collected at several time points (1, 3, 5, 10 and 24 h) after stress initiation. Samples were immediately frozen in liquid nitrogen and stored at -80 °C prior to RNA extraction and qRT-PCR analysis. Four independent biological replicates were performed for MeJA, SA and NaCl stress, two biological replicates were performed for ABA and heat stress, all with two technical replicates.

3.3.5 Biotic stress treatments

For the *Pseudomonas syringae* and *Botrytis cinerea* experiment, individually grown 5-week-old WT *Arabidopsis* plants of the Col-0 ecotype sown in Jiffy's were used. These plants were maintained in a controlled growth chamber (Convion) at 21 °C and 12/12 h light/dark photoperiod. The *Pseudomonas syringae* pv. *tomato* DC3000 strain and the *Botrytis cinerea* B05.10 strain were supplied by the Phytopathology lab of Prof. dr. M. Höfte (Ghent University, Belgium). Infection assays were performed according to Pieterse *et al.* (1996), Audenaert *et al.* (2002) and Katagiri *et al.* (2002), with some minor modifications.

P. syringae pv. *tomato* DC3000 was grown in liquid King's B medium (20 g/L peptone, 1 % glycerol, 1.5 g/L KH_2PO_4 , 1.5 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, pH 7.2) at 28 °C on a rotary shaker (200 rpm) until the culture reached the mid to late log growth phase ($\text{OD}_{600} = 0.6-1.0$). After centrifugation of the culture (10 min, 2500 g), bacterial cells were resuspended in 10 mM MgSO_4 to obtain a solution of bacteria with an OD_{600} of 0.05 (corresponding to 2.5×10^7 cfu/mL). Prior to use 0.05 % Silwet-77 (GE Specialty Materials, Switzerland) was added to the infection solution. The mock solution consisted of 10 mM MgSO_4 containing 0.05 % Silwet-77, without bacteria. The rosette leaves of the *Arabidopsis* plants were sprayed until run-off with either the infection or the mock solution. To increase the efficiency of the infection, the plants were maintained at 100 % relative humidity one day before the treatment until two days after the start of the bacterial infection.

The *Botrytis* strain was kept on regular potato dextrose agar plates at 21 °C. Sporulation was stimulated by incubation for 10 days at 21 °C under a 12/12 h UV/dark light regime (UV-A: 365 nm; combination normal TL and TL black light blue lamp). After 10 days, *Botrytis* spores were harvested by washing the plates with distilled water containing 0.01 % Tween-20 (VWR). This suspension was filtered through a nylon membrane (20 μm \varnothing) and the conidia were counted using a Bürker counting chamber. The inoculation suspension (5×10^5 conidia/mL) was prepared in 1/2 strength potato dextrose broth medium. The mock solution consisted of the same medium without spores. The droplet technique was used for the infection: a 10 μL droplet of either the infection suspension or the mock solution was added on the upper side of three rosette leaves from each plant. 100 % relative humidity was maintained during the entire experiment.

During both infection assays, plants were kept in a controlled Conviron growth chamber at 21 °C with a 12/12 h light/dark photoperiod. Control plants were kept separately from infected plants. Rosette leaves of 8-10 randomly chosen plants were sampled in liquid nitrogen at different time points post-infection and samples were stored at -80 °C prior to RNA extraction. In case of the *P. syringae* infection experiment, three biological replicates were performed and analyzed, for the *B. cinerea* infection experiment, two biological replicates were performed and analyzed, all with two technical replicates.

Myzus persicae was kindly provided by the Agrozoology lab of Prof. dr. Guy Smagghe (University Ghent, Belgium) and kept on sweet pepper plants under lab conditions (Shahidi-Noghabi *et al.*, 2009). Aphid infestation was performed on 5-week-old *Arabidopsis* plants sown in round plastic pots (\varnothing 11 cm) with soil on which a transparent ventilated cage (Novolab) was placed. About 60 aphids (adults) were placed on the rosette leaves of each plant. After putting the aphids, the cages around the separate plants were closed. Control plants were grown in a cage without aphids. During the assay, plants were kept in a controlled Conviron growth chamber at 21 °C with a 12/12 h light/dark photoperiod. At indicated time points, aphids from two leaves of nine randomly chosen plants were removed and leaves were sampled and frozen in liquid nitrogen. Samples were stored at -80 °C prior

to RNA extraction and qRT-PCR analysis. Four biological replicates were performed and analyzed, each with two technical replicates.

3.3.6 RNA extraction, cDNA synthesis and RT-PCR analysis

All plant samples were first homogenized using a mortar and a pestle, and total RNA was extracted using TRI Reagent[®] according to the instructions of the manufacturer (Sigma-Aldrich). The RNA samples were treated with DNase I (Life Technologies) to remove the residual genomic DNA. Shortly, 12 μ L RNA was incubated for 30 minutes at 37 °C with 2 μ L of RNase-free DNase I, 2 μ L DNase I buffer (10x) and 4 μ L of distilled water. After addition of 2 μ L 25 mM EDTA, the reaction was inactivated for 10 minutes at 65 °C. The RNA concentration was measured with a Nanodrop 2000 spectrophotometer (Thermo Scientific). cDNA was synthesized from the RNA using the moloney murine leukemia virus (M-MLV) transcriptase kit (Life Technologies). Briefly, 1 μ g of DNase treated RNA was incubated for 5 min at 65 °C with 1 μ L 2 μ M oligo(dT)₂₅ primer, 1 μ L 10 mM dNTPs and distilled water to a total volume of 12 μ L. Then, 4 μ L of M-MLV buffer and 2 μ L of dithiothreitol (0.1 M) were added and this mixture was incubated for 2 min at 37 °C. Finally 1 μ L M-MLV RT was added and an incubation of 50 min at 37 °C and 15 min at 75 °C was executed.

The full length cDNA sequences of *AN3*, *AN4* and *AN5* were retrieved by RT-PCR reactions with gene specific primers (Supplementary table 4). The PCR reaction mixture was as follows: 2 μ L cDNA, 2 μ L 10 mM dNTPs (Thermo Fisher Scientific), 2.5 μ L 10 x RxN buffer (VWR), 1 μ L 5 μ M forward and 1 μ L 5 μ M reverse primer (Life Technologies), 0.75 μ L 50 mM MgCl₂, 0.125 μ L Platinum[®] Pfx DNA Polymerase (Life Technologies) and water up to the volume of 25 μ L. The PCR conditions used were: 2' 95 °C - 30-35 x (15'' 94 °C - 30'' 47-50 °C - 1' 72 °C) - 5' 72 °C. After cloning these sequences in the pJET2.1 vector with the CloneJET PCR Cloning kit (Life Technologies), the constructs were checked by agarose gel electrophoresis and sequenced (LGC Genomics, Berlin, Germany) to confirm the correct cDNA sequence of the *Nictaba* homologs from *A. thaliana*.

Before performing RT-PCR, cDNA quality was checked by RT-PCR using primers specific for the protein phosphatase 2A (PP2A) gene (Supplementary table 5). The PCR reaction mixture is the same as previously mentioned except for the buffer (10 x EXTRA buffer, VWR) and the enzyme (*Taq* DNA polymerase, VWR). The PCR conditions are as follows: 5' 95 °C – 45 x (45'' 94 °C - 45'' 55 °C - 30'' 72 °C) - 5' 72 °C. PCR amplification products were checked by agarose gel electrophoresis (2.5 % agarose gel in 0.5 x Tris-acetate-EDTA (TAE) buffer).

3.3.7 Construction of EGFP fusion constructs

Coding sequences for the *Nictaba* homologs from *A. thaliana* were N- and C-terminally fused to EGFP using the Gateway[®] Cloning Technology (Life Technologies, Carlsbad, CA, USA). The cloned full length cDNA sequences (in the pJET1.2 vector) were used as a template to amplify the open reading frames with primers to attach *attB* sites. In the first PCR, the first part of the *attB* site is attached by amplifying the open reading frame using Platinum[®] Pfx

DNA Polymerase (Life Technologies) and primers with a gene specific part (with or without stop codon) and the first part of the *attB* site (Supplementary table 6). In the second PCR, a 1:5 dilution of the first PCR product was used as a template in combination with primers to complete the *attB* sites (Supplementary table 6). The PCR conditions for the first PCR were as follows: 2' 94 °C – 30 x (15" 94 °C - 30" 50 °C - 1' 72 °C) - 5' 72 °C. The conditions for the second PCR were: 2' 94 °C – 5 x (15" 94 °C - 30" 48 °C - 1' 72 °C) – 25 x (15" 94 °C - 30" 55 °C - 1' 72 °C) - 5' 72 °C. After checking the PCR products by agarose gel electrophoresis, the PCR fragments were used in a BP recombination reaction with the pDONR221 vector. The *attB* PCR products and the pDONR221 vector were incubated overnight in equimolar amounts with the BP Clonase® II enzyme mix. The next day, the resulting entry clones (with and without stop codon) were transformed into heat-shock competent *E. coli* cells (strain TOP 10) and transformants were selected on LB agar plates with 50 µg/mL kanamycin. Subsequently, transformants were checked with colony PCR and agarose gel electrophoresis. The transformants containing an entry clone from the expected size were grown and the entry clones were extracted using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The sequences of the entry clones were checked by LGC Genomics (Berlin, Germany) and the correct sequences were used in a LR recombination reaction with the desired destination vectors (containing a CaMV 35S promoter and the EGFP gene). Entry clones of the Nictaba homolog open reading frame with stop codon were used to make the N-terminal EGFP fusions with the pK7WGF2,0 destination vector (Karimi *et al.*, 2002). Entry clones of the Nictaba homolog open reading frame without stop codon were used to make the C-terminal EGFP fusions with the pK7FWG2,0 destination vector (Karimi *et al.*, 2002). Similarly, this recombination reaction was incubated overnight according to the Gateway® manual and expression clones were transformed via heat shock into *E. coli* TOP10 cells. Transformants were selected on LB agar plates with 75 µg/mL spectinomycin and screened with colony PCR using gene specific and EGFP primers (Supplementary table 6).

3.3.8 Transient transformation of *N. benthamiana* leaves

The expression vectors containing the different EGFP fusion constructs were introduced into *A. tumefaciens* C58C1 pMP90 cells using triparental mating. Briefly, a donor strain (*E. coli* containing the expression vectors), a helper strain and the Agrobacterium are mixed together on solid YEB medium (5 g/L beef extract, 5 g/L peptone, 1 g/L yeast extract, 5 g/L sucrose and 15 g/L bacterial agar) containing 2 mM MgSO₄. The strains have to interact with each other to introduce the expression vectors in Agrobacterium. After incubation, a dilution series was made and transformants were selected on solid YEB medium containing 75 µg/mL spectinomycin and 20 µg/mL gentamycin. Screening of the colonies was done by PCR with the same primers used for the LR colony PCR (Supplementary table 6) after purification of the expression clones out of the Agrobacterium.

Transient expression of the EGFP fusion constructs was obtained by infiltration of the transformed *Agrobacterium* in leaves of 4- to 6-week-old *N. benthamiana* plants as described by Sparkes *et al.* (2006). First, the *Agrobacterium* strains were grown in liquid YEB medium containing 75 µg/mL spectinomycin and 20 µg/mL gentamycin for two days at 25 °C on a rotary shaker (200 rpm). *Agrobacterium* cells were harvested by centrifugation and resuspended in infiltration medium (50 mM MES, 2 mM Na₂HPO₄, 0.5 % glucose, pH 5.6). Centrifugation and resuspension was repeated twice, the second time using infiltration medium with 100 µM acetosyringone. After washing, the cells were diluted to a final optical density at 600 nm of 0.01, 0.05, 0.1 and 0.2 and infiltrated in the leaf epidermal cells. The spot of infiltration was labelled and plants were further grown in the growth chamber. Two or three days post-infiltration, microscopic analysis and Western blot analysis were performed.

3.3.9 Stable transformation of *A. thaliana* plants

Stably transformed *Arabidopsis* plants were created using the floral dip transformation method (Clough and Bent, 1998). The inflorescences of approximately 7-week-old plants grown in pots with commercial soil were cut. After one week, new and more shoots were formed and these plants were used for transformation. Transformed seeds were selected on MS medium containing 75 µg/mL kanamycin (Duchefa) using the fast selection protocol according to Harrison *et al.* (2006). Green plantlets were transferred to new selective MS medium and afterwards to Jiffy's. Integration of the T-DNA was checked by PCR on gDNA using gene specific and EGFP primers (Supplementary table 6). T2 generation *Arabidopsis* plants were used for all analyses.

3.3.10 Confocal microscopy and image analysis

Microscopic analysis was performed using the confocal laser scanning microscope Nikon A1R (Nikon Belux). The lower epidermis of the leaf discs (spots that are infiltrated) were visualized using the 40 x S Plan Fluor ELWD air objective lens (NA of 0.6). All images are a combination of different fluorescent images acquired along the z-axis, as such the complete epidermis cell could be visualized. EGFP was excited with a 488 nm argon ion laser, fluorescent emission filters were 525 nm for EGFP (with band width of 50 nm) and 700 nm for autofluorescence of chlorophyll (with band width of 75 nm). All images were created by the software package NIS-Elements (Nikon) and image analysis was performed using Fiji (Schindelin *et al.*, 2012).

3.3.11 qRT-PCR analysis

Before doing qRT-PCR, cDNA quality was checked by RT-PCR using primers specific for the *PP2A* gene (Supplementary table 5) as described in section 3.3.6. Real time qRT-PCR analyses of the gene expression during development and after *P. syringae* infection, were performed using the Rotor-Gene 3000 (Corbet Life Science) and the Rotor Discs (Qiagen, Hilden, Germany). The program was as follows: 10' 95 °C – 45 x (25" 95 °C - 25" 60 °C - 20" 72 °C) -

5' 72 °C followed by generation of a melting curve (gradual increase from 72 °C to 95 °C with 1 °C/step). The Rotor Gene 6 software generated the raw output data (Cq values), these were statistically analyzed using the REST-384 software (Corbett Research). REST-384 uses a pair wise fixed reallocation randomization test as a statistical model (Pfaffl *et al.*, 2002). Three independent biological replicates of the infection experiment with *P. syringae* were performed, two biological replicates of the gene expression analysis experiment during development of the plant, each with two technical replicates.

qRT-PCR analyses of the other stress experiments were performed using the 96-well CFX Connect™ Real-Time PCR Detection System (Bio-Rad). The program was as follows: 10' 95 °C - 45 x (15" 95 °C - 25" 60 °C - 20" 72 °C) followed by generation of a melting curve (gradual increase from 65 °C to 95 °C with 0.5 °C/step). The CFX Manager 3.1 software (Bio-Rad) generated the raw output data which were statistically analyzed using the REST-384 software. Two or four independent biological replicates were performed with each two technical replicates and analyzed together using the sample maximization approach (Hellemans *et al.*, 2007).

All reactions were conducted in a total volume of 20 µL containing 1 x SensiMix™ SYBR® No-ROX One-Step mix, 2 µL undiluted cDNA template, 500 nM gene specific forward and reverse primers (Supplementary table 5). All gene specific qPCR primers were designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). Specificity of the primers was tested *in silico* by BLAST search and amplicons were cloned and verified using agarose gel electrophoresis and sequencing (LGC Genomics, Berlin). Amplification efficiency of all primer pairs was determined in the CFX Manager 3.1 (Bio-Rad) and qBASE^{PLUS} software (Hellemans *et al.*, 2007). All expression data were normalized using three reference genes: *PP2A*, *TIP41* and *UBC9* (Czechowski *et al.*, 2005). All melting curves were checked and reference gene stability and quality control of the samples were validated in the qBASE^{PLUS} software (Hellemans *et al.*, 2007).

3.4 Results

3.4.1 Nictaba homologs in *A. thaliana*

The genome of *A. thaliana* contains 30 sequences encoding ArathNictabas, identified in chapter 2. A large group of these Nictaba homologs (19/30) have an N-terminal F-box domain, next to their carbohydrate binding domain. Furthermore, four Nictaba homologs were identified with an N-terminal TIR domain and one with an N-terminal AIG1-type G domain (Eggermont *et al.*, 2017; Chapter 2). This study focuses on the non-chimeric Nictaba homologs from *A. thaliana* and we want to investigate whether these Nictaba homologs have similar biological properties as the lectin from tobacco.

Based on protein sequence alignments with Nictaba, the length of the N-terminal unknown domains and identified ESTs, five cDNA sequences of non-chimeric Nictaba homologs were initially selected for further research (Table 3.1). Coding sequences for each ArathNictaba (AN) were amplified from leaf cDNA, cloned and used to make different expression constructs. The subcellular localization of EGFP fusion constructs for the ArathNictabas was investigated in transiently transformed *N. benthamiana* leaves and stably transformed *A. thaliana* plants. Three out of five ArathNictabas were selected to unravel their biological relevance since the EGFP fusion constructs of AN1 and AN2 did not show any fluorescence. These three ArathNictabas are further specified as ArathNictaba3 (AN3; AT4G19850.2), ArathNictaba4 (AN4; AT1G31200) and ArathNictaba5 (AN5; AT4G19840).

Table 3.1 Five selected non-chimeric Nictaba homologs.

Non-chimeric Nictaba homologs	AGI code
AN1	AT1g33920
AN2	AT4G19850.1
AN3	AT4G19850.2
AN4	AT1G31200
AN5	AT4G19840

The ArathNictabas contain next to their Nictaba domain, an N-terminal sequence of different lengths (Figure 3.1). In AN3, AN4 and AN5, the Nictaba domain is preceded by an N-terminal sequence of 61, 28 and 91 aa, respectively. Protein BLAST with these N-terminal parts concedes no homology to any other plant protein domain. The length of the Nictaba domain is similar in all three ArathNictabas, in particular 135, 145 and 154 aa for AN3, AN4 and AN5, respectively. The total molecular weight and the pI of each ArathNictaba are indicated in Figure 3.1. A search for signal peptides and transmembrane regions using the Phobius and SignalP 4.1 server (Käll *et al.*, 2004; Petersen *et al.*, 2011) revealed no signal peptides or transmembrane regions in the sequences of the ArathNictabas. Judging from the absence of signal peptides, AN3, AN4 and AN5 are probably synthesized on free ribosomes in the cytoplasm. Moreover, no evidence was found for a nuclear localization of these proteins using the NucPred server (Brameier *et al.*, 2007). The cytosol is also the predicted localization for these three ArathNictabas according to the SUBA3 server (Hooper *et al.*, 2014).



Figure 3.1 Domain architectures of the ArathNictabas. The domain architectures are drawn to scale. The scale bar represents 100 aa.

3.4.2 Sequence similarity between the Nictaba homologs from *A. thaliana* and Nictaba

Protein sequence alignment of the Nictaba domains for the ArathNictabas and Nictaba allowed to calculate the sequence identities and similarities (Figure 3.2). The percentages of sequence identity and sequence similarity between Nictaba and the ArathNictabas are indicated in Table 3.2 and Table 3.3. The highest percentage of sequence identity is found between AN3 and AN5. Moreover, their sequence identity with Nictaba is very similar, 31.8 % and 31.5 % respectively. The lowest percentage of sequence identity with Nictaba is observed for AN4 (22.9 %). The putative NLS reported for the Nictaba sequence from tobacco (K102-K105) is not conserved in the ArathNictaba sequences. The tryptophan residues important for carbohydrate binding in Nictaba are conserved in all ArathNictaba sequences (Figure 3.2). The glutamic acid residues that are conserved in a lot of sequences encoding Nictaba homologs, are partly conserved in the ArathNictaba sequences. Although these glutamic acid residues are less conserved, nearby glutamic acid residues probably retain the electronegative character of the presumed carbohydrate binding site (Figure 3.2).

Table 3.2 Protein sequence identity between Nictaba and ArathNictaba domain sequences.

%	Nictaba	AN3	AN4	AN5
Nictaba	100	31.8	22.9	31.5
AN3	31.8	100	26.0	48.4
AN4	22.9	26.0	100	28.7
AN5	31.5	48.4	28.7	100

Table 3.3 Protein sequence similarity between Nictaba and ArathNictaba domain sequences.

%	Nictaba	AN3	AN4	AN5
Nictaba	100	45.7	39.4	46.1
AN3	45.7	100	43.5	67.5
AN4	39.4	43.5	100	38.9
AN5	46.1	67.5	38.9	100

```

AN4          -----SGLNFVWGGD--SRYWVIPKEP-----RMPAELKMVSWLEVTGSFD--K
Nictaba      MQGQWIAARDLSITWVDN--PQYWTWTKTV-----DPNIEVAELRRVAWLDIYGKIETKN
AN3          KNCFMLYARDLSITWAESQTNKYWSWFSDLDQTSDDVRTEVAKMERVAWLEVVGKFETEK
AN5          SNCFMLFAKNLSITWSDD--VNYWTWTFTEKESPEN--VEAVGLKNVCWLDITGKFDTRN
              .*.:.* . .** . . . . *.**: : *.: : :

AN4          IEPGKTYRIGFKISFKPDATGWDKAPVFMSAK-----IGKKGKTVWKRIKSVSQNFGIL
Nictaba      LIRKTSYAVYLVFKLTDNPRELERATASLRFVNEVAEGAGIEGTTVFISKKK-----
AN3          LTPNSLYEVVVFVVKLIDSAKGWDFRVN-FKLV-----LPTGET-----
AN5          LTPGIVYEVVFKVKLEDPAYGWDTPVN-LKLV-----LPNGKE-----
              :      * : : .: :      :      :      :

AN4          KGGSEPVNIP-DESDGLFEILVSP---TALNQDTKLQFGLYEVWTGRWKTGLLIH-----
Nictaba      KLPGELGRFPHLRSDGWLEIKLGEFFNNLGE-DGEVEMRLMEINDKTWKSIIVKGFDIR
AN3          KERRENV--NLLERNKWVEIPAGEFMISPEHLSGKIEIRK-----
AN5          KPQEKVSLRELPRYKWDVVRVGEFVPEKSA-AGEITFSMYEHAAGVWKKGLSLKGVAIR
              *      :      .: : .      : : :

AN4          --
Nictaba      PN
AN3          --
AN5          PK

```

Figure 3.2 Protein sequence alignment of the Nictaba sequence and the Nictaba domains of AN3, AN4 and AN5 from *A. thaliana* using Clustal Omega. The tryptophan residues important for carbohydrate binding in Nictaba are marked in bold (Trp15, Trp22). The glutamic acid residues (Glu138 and Glu145) are marked in bold and underlined. The NLS of Nictaba is underlined.

3.4.3 The Nictaba homologs from Arabidopsis show a nucleocytoplasmic localization

To validate the *in silico* results (section 3.4.1), N- and C-terminal EGFP fusion constructs under a 35S promoter were created using the Gateway cloning system. Leaves from *N. benthamiana* were transiently transformed with the N- and C-terminal EGFP fusion constructs for each ArathNictaba sequence. Figure 3.3A shows a nucleocytoplasmic localization for AN4 and AN5 whereas AN3 only resides in the cytoplasm. The localization of AN4 and AN5 into the cytoplasm and the nucleus is confirmed for both N- and C-terminal EGFP fusion constructs. The N-terminal EGFP fusion construct of AN3 never showed a fluorescence signal. The highly fluorescent dots observed for the C-terminal EGFP fusion construct of AN3 were not observed in all pictures, and therefore can probably be considered as an artifact. Figure 3.3B confirms the fluorescence within the nucleus for all ArathNictaba EGFP fusion constructs except for AN3-EGFP where the fluorescence in the nucleus is almost negligible. The fluorescence of EGFP-AN4 is high in the nucleus compared to the surrounding cell compartments whereas the fluorescence of AN4-EGFP is almost the same in the nucleus as in the surrounding cell compartments, suggesting that the transport of AN4-EGFP to the nucleus is much less efficient than the transport of EGFP-AN4 into the nucleus. The fluorescence of both EGFP-AN5 and AN5-EGFP is comparable and the signal in the nucleus is high compared to the surrounding cell compartments.

Microscopic analysis of the T2 generation of Arabidopsis plants stably transformed with the EGFP fusion constructs confirmed the results obtained with *N. benthamiana* (Figure 3.4). In both experiments, free EGFP was used as a positive control and is observed in the nucleus

and the cytoplasm (data not shown). EGFP has a molecular weight of 26.9 kDa and as such can freely diffuse from the cytoplasm into the nucleus.

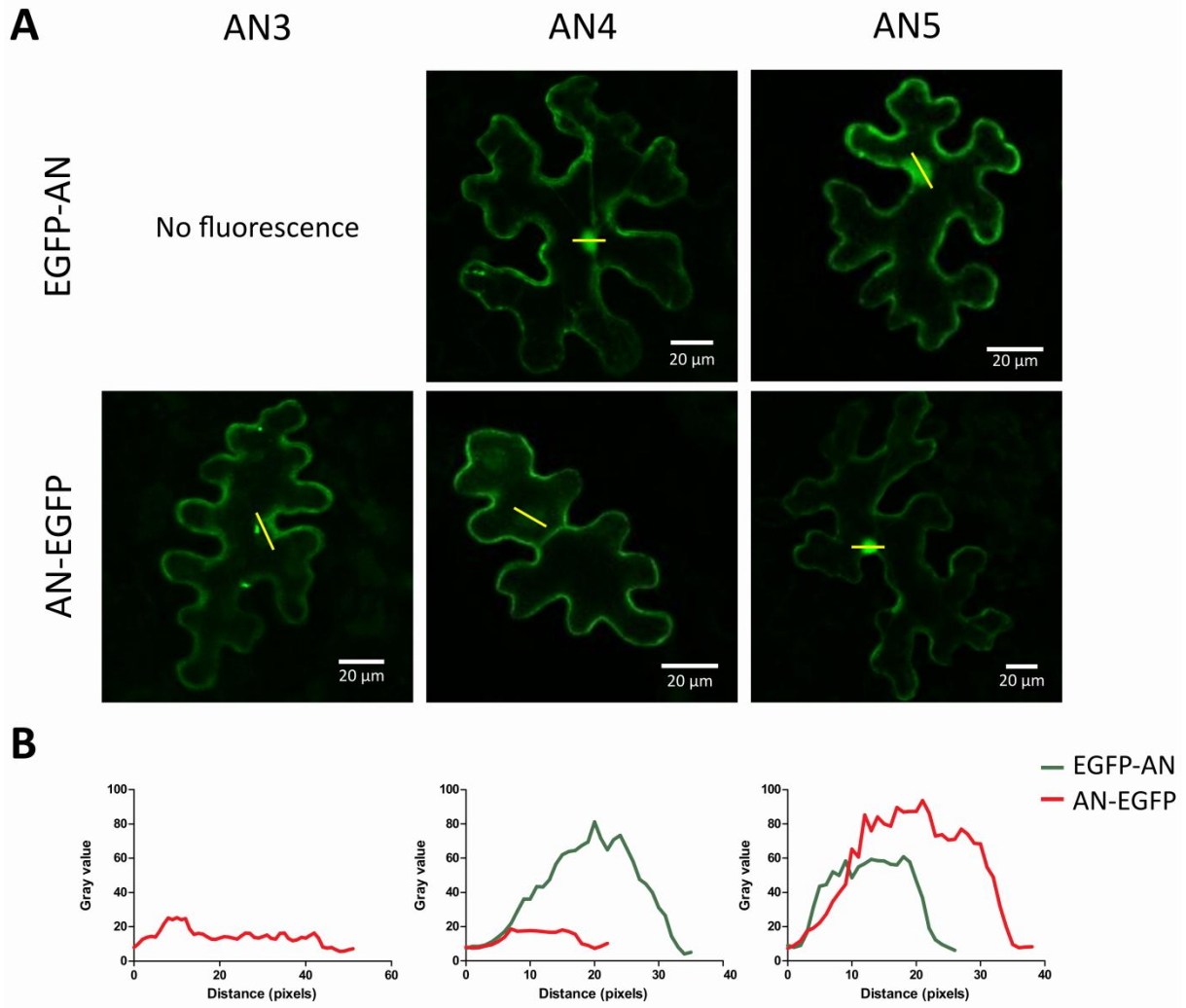


Figure 3.3 Localization of N- and C-terminal EGFP fusion constructs of ArathNictabas expressed in transiently transformed *N. benthamiana* leaves. Localization is shown in the lower epidermis cells of the leaves. (A) Confocal microscopy images of all EGFP fusion constructs of the ArathNictabas. All images are a compilation of different fluorescent images acquired along the z-axis. (B) Plot profiles showing the fluorescence intensity within the nucleus (yellow line in panel A).

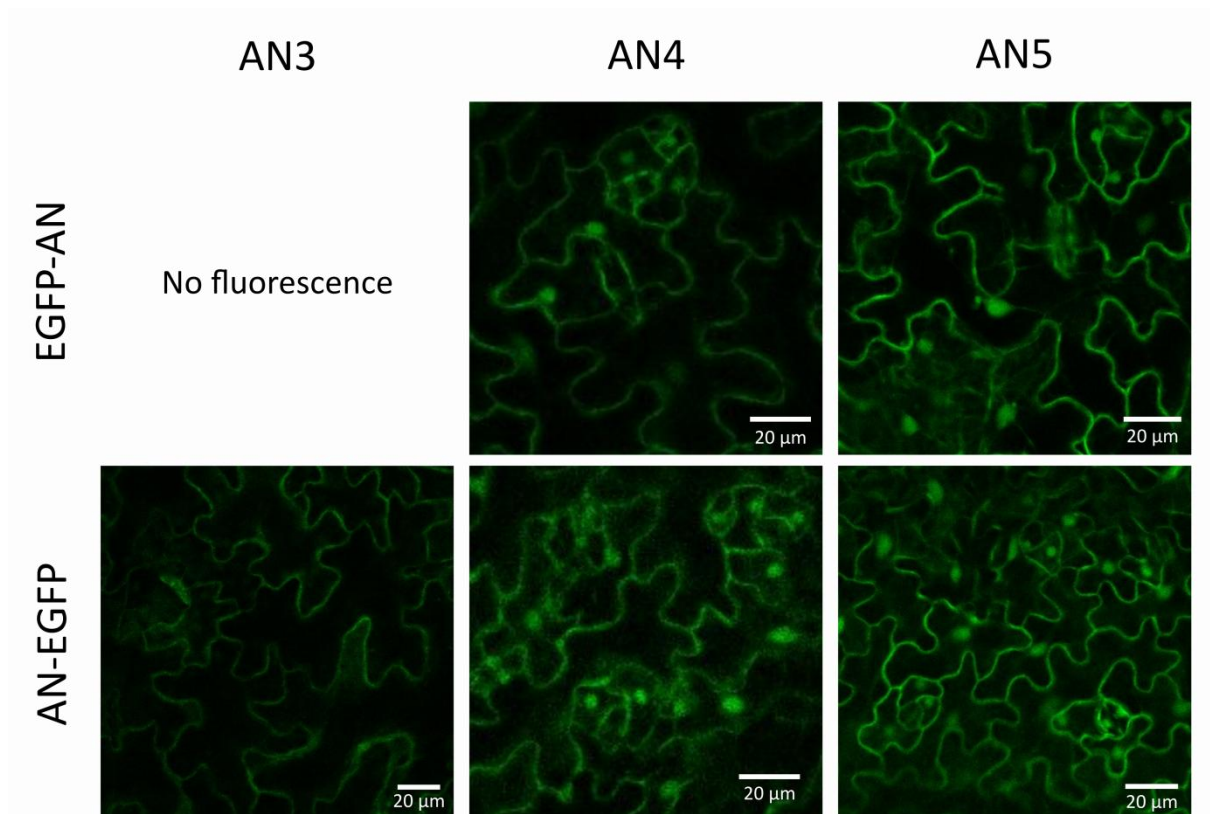


Figure 3.4 Localization of N- and C-terminal EGFP fusion constructs of *ArathNictabas* expressed in stably transformed *A. thaliana* plants. Localization is shown in the lower epidermis cells of the leaves. All images are a compilation of different fluorescent images acquired along the z-axis.

3.4.4 *In silico* expression analysis of the *Nictaba* homologs from *A. thaliana*

An *in silico* analysis of the 1500 bp predicted promoter region of *AN3*, *AN4* and *AN5* was performed. PlantCARE (Lescot *et al.*, 2002) and AGRIS (Yilmaz *et al.*, 2011) were used to search for experimentally validated and predicted cis-regulatory elements and TF binding sites in the different promoter regions of the *ArathNictaba* sequences. The CAAT and TATA box, the core promoter elements, were found in each promoter sequence, indicating that the sequences encode a functional promoter (data not shown). Next to these core promoter elements, several protein binding sites, tissue-specific and stress responsive elements were found in the promoters of *AN3*, *AN4* and *AN5* (Figure 3.5). Light-responsive elements constitute the largest group in the three promoter sequences, these elements are not indicated in Figure 3.5. The promoter of *AN3* is the only one not containing MeJA responsive elements. All promoters contain SA responsive elements. Interestingly, the promoter of *AN4* contains a root meristem specific element and the promoter of *AN5* a general meristem specific element, next to a few shoot specific elements. The promoter of *AN3* contains a heat responsive element whereas the promoter of *AN4* contains several osmotic stress responsive and ET responsive elements. The promoter of *AN5* contains a drought responsive element. The presence of all these stress responsive cis-regulatory elements suggests that *ArathNictabas* may play a role in the stress response of the plant.

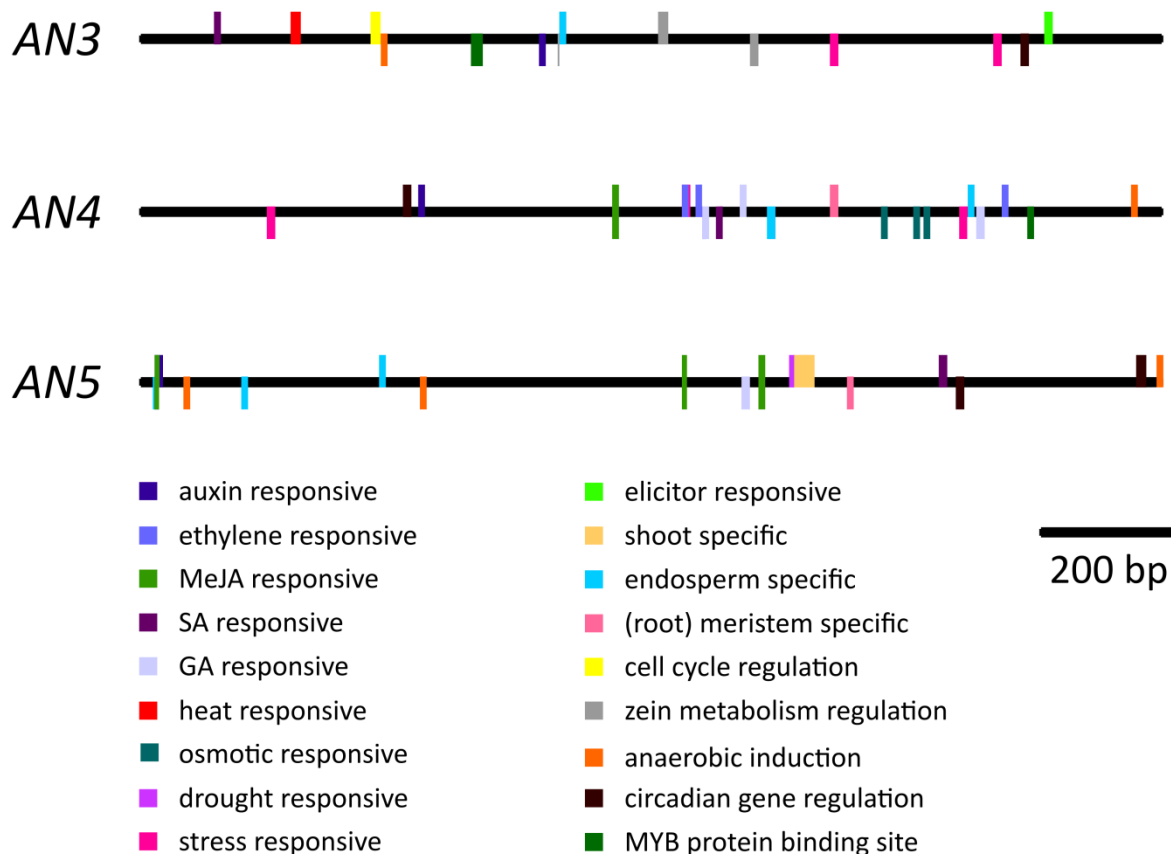


Figure 3.5 Schematic representation of the 1500 bp promoter sequences for AN3, AN4 and AN5. Protein binding sites, tissue-specific and stress responsive cis-regulatory elements are shown in different colors. CAAT, TATA boxes and light responsive elements were not included. Cis-regulatory elements present in the sense and antisense strand are shown on top or below the promoter sequences, respectively.

In a second approach to obtain information about the expression of the *ArathNictabas*, BLASTn and tBLASTn searches were performed against the EST database of NCBI. This analysis revealed evidence for expression of AN3, AN4 and AN5 in the aerial tissues of 8-day-old *A. thaliana* plants grown in continuous light (Weber *et al.*, 2007) and in pooled cDNA of *Arabidopsis* roots, inflorescence, callus, young seedlings and *Arabidopsis* treated with cold, heat, salt, hydrogen peroxide, UV, auxin, *Xanthomonas* and *Pseudomonas*. Furthermore, ESTs for AN3 were identified in the green siliques and aboveground organs of two- to six-week-old plants (Asamizu *et al.*, 2000). ESTs for AN4 and AN5 revealed evidence for expression of AN4 and AN5 in the roots of *A. thaliana* (Asamizu *et al.*, 2000). Moreover, AN5 ESTs were also identified in plants one week after bolting (Feilner *et al.*, 2005), in plants at various developmental stages subjected to dehydration and cold stress, in the aboveground organs of two- to six-week-old plants (Asamizu *et al.*, 2000), floral buds (Alexandrov *et al.*, 2006), green siliques (Asamizu *et al.*, 2000) and in adult vegetative tissues. As illustrated in this EST analysis, stress-related ESTs were identified for all *ArathNictabas*, suggesting a possible role in the plant stress response.

The eFP browser contains gene expression data from the ATH1 GeneChip from Affymetrix and the AtGenExpress initiative (Winter *et al.*, 2007). Genevestigator is structuring data from Affymetrix expression microarrays (Hruz *et al.*, 2008). Screening of this data revealed more information on the tissue specific expression and expression upon different stress treatments for the three *ArathNictabas*. According to these data, the overall expression level of *AN3* is low and *AN3* is mostly expressed in the seeds, stems, flowers and pollen. The overall expression level of *AN4* is also low and is significantly higher in the roots, stems, leaves and pollen. The expression of *AN5* is relatively higher than the expression of *AN3* and *AN4*, and *AN5* is mostly expressed in the rosette and cauline leaves, hypocotyl, stems and flowers.

The expression upon different stresses is illustrated in a heat map (Figure 3.6). The expression level of *AN3* is highly influenced (up- and downregulation) by different stress factors, while the changes in expression level for *AN4* are only minor after different stress treatments. The changes in expression level for *AN5* after different stress treatments are almost negligible.

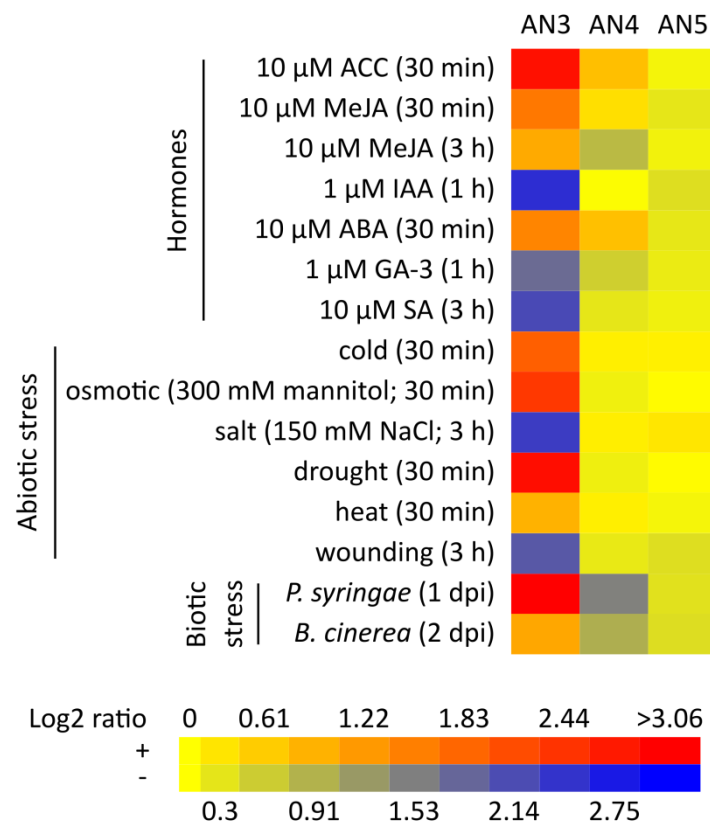


Figure 3.6 Relative expression levels of *ArathNictabas* after several stress treatments inferred from microarray data. Log₂ transformed microarray data (Winter *et al.*, 2007) visualized in a heat map created by the BAR HeatMapper Plus Tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi). ACC: 1-aminocyclopropane-1-carboxylic acid, IAA: indole-3-acetic acid, GA: gibberellic acid.

3.4.5 Expression of the *ArathNictaba* genes during development of WT *A. thaliana* plants

Using qRT-PCR, the expression level of the *ArathNictabas* was investigated in different tissues from *Arabidopsis* during development. *Arabidopsis* plants were grown under standard conditions and plant material was collected at different developmental stages starting from 6-day-old plantlets to different tissues of 54-day-old plants. Figure 3.8 shows the normalized relative expression for the three *ArathNictabas* throughout the development of the plant relative to the expression of these *ArathNictabas* in 6-day-old plantlets (first developmental stage). The three *ArathNictaba* genes are expressed in every tissue during all developmental stages tested. The expression level of *AN3* is significantly higher in the stems and the flowers compared to the expression in 6-day-old plantlets. The expression level of *AN4* is significantly higher in the roots and significantly lower in the flowers of the plant compared to its expression in 6-day-old plantlets. The expression level of *AN5* is significantly higher in the rosette- and cauline leaves at all developmental stages tested. Moreover, the expression level of *AN5* is significantly lower in the flowers compared to its expression in 6-day-old plantlets.

The normalized expression of *AN3* and *AN4* compared to the expression of one of the three reference genes (*PP2A*, *TIP41* or *UBC9*) is much lower than the normalized expression of *AN5* compared to the expression of the same reference gene, indicating that the expression level of *AN5* is higher than the expression levels of *AN3* and *AN4* in all tissues throughout the development of the plant (Figure 3.7).

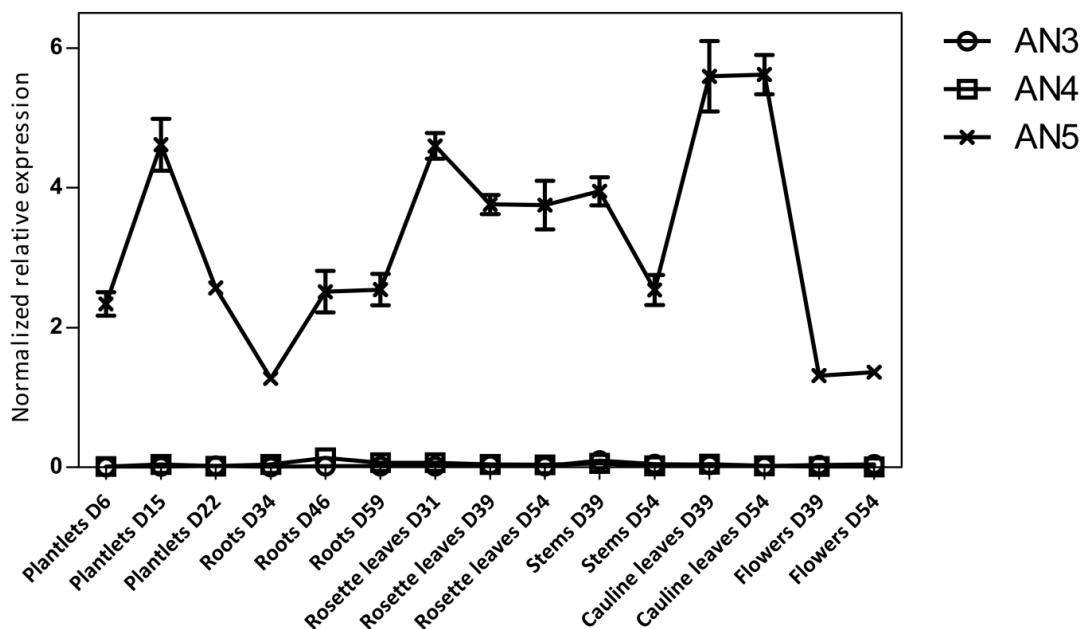


Figure 3.7 Normalized expression of the three *ArathNictaba* genes relative to the expression of *TIP41* during the development of *A. thaliana*. The normalized transcript levels (to three reference genes: *PP2A*, *TIP41* and *UBC9*) are the result of two independent biological replicates (N = 2) with similar results.

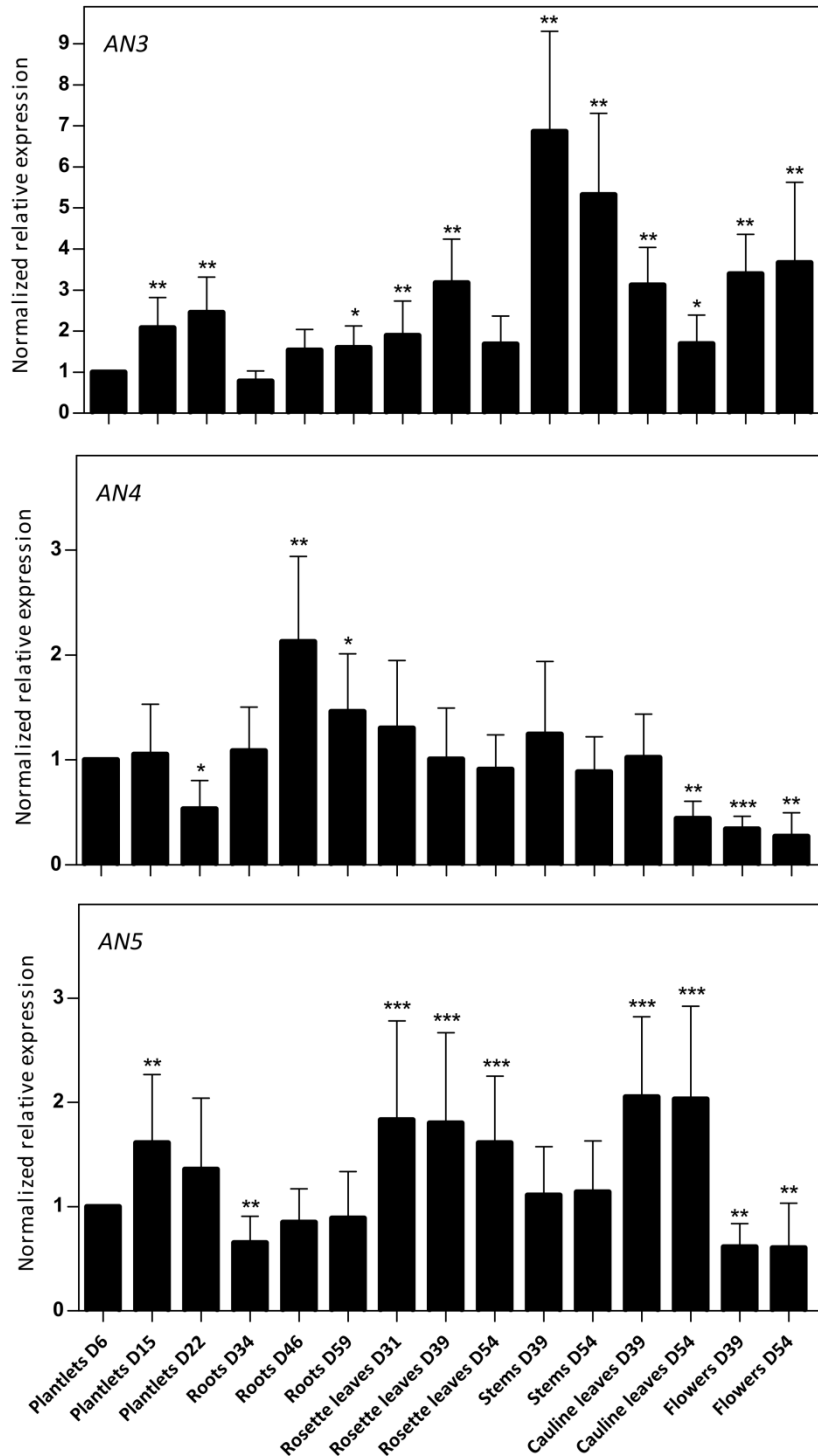


Figure 3.8 Normalized expression of the three *ArathNictaba* genes relative to the *ArathNictaba* expression level in 6-day-old (D6) plantlets during the development of *A. thaliana*. The normalized transcript levels (to three reference genes: *PP2A*, *TIP41* and *UBC9*) are the result of two independent biological replicates (N = 2) with similar results. Bars represent the mean \pm SE normalized relative expression and asterisks indicate statistically significant differences to the expression level of *ArathNictaba* in 6-day-old plantlets (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; REST analysis).

3.4.6 *ArathNictaba* expression is stress-inducible

Judging from the *in silico* expression analysis, several hormones, abiotic and biotic stresses can influence the expression of the *ArathNictabas*. Based on the data from eFP browser, Genevestigator and recent research on other Nictaba homologs in our research group, several stress conditions were selected in order to get a profound insight in the stress inducibility of the *ArathNictaba* expression.

3.4.6.1 *AN3, AN4 and AN5 are differentially expressed in response to hormone treatments*

Sixteen-day-old *Arabidopsis* seedlings were subjected to different hormone treatments in particular 100 μ M MeJA, 100 μ M ABA and 300 μ M SA. Mock/hormone treated samples were collected 1, 3, 5, 10 and 24 hours after treatment and transcript levels for *AN3*, *AN4* and *AN5* were determined by qPCR analysis (Figure 3.9). Control genes known to be responsive to MeJA (*JMT*), ABA (*Cor15A*) and SA (*WRKY70*) treatments are significantly upregulated indicating that the plants sensed the different stress treatments.

The expression of *AN3* is significantly upregulated after treatment with different hormones with a fourfold upregulation after 5, 10 and 24 hours of MeJA treatment. After ABA and SA treatment, the upregulation of the expression of *AN3* is less pronounced. ABA treatment resulted in a twofold upregulation after 5 and 10 hours, SA treatment in a 1.6 - 3.6 fold upregulation after 1, 3, 5, 10 and 24 hours.

In contrast with the expression of *AN3*, the expression of *AN4* is mostly downregulated after treatment with different hormones with a 2 - 2.5 fold downregulation after 5, 10 and 24 hours of MeJA treatment. Whereas there was a twofold upregulation of the expression of *AN3* after 5 and 10 hours of ABA treatment, the expression of *AN4* is approximately two times downregulated. After SA treatment, the expression of *AN4* is not changed except for a small significant downregulation after 10 hours.

The expression of *AN5* is only weakly influenced by the MeJA treatment. Similar to the expression of *AN3*, there is a two times upregulation of the expression of *AN5* after 5 and 10 hours of ABA treatment. After SA treatment, the expression of *AN5* is not changed except for a small significant upregulation after 10 hours.

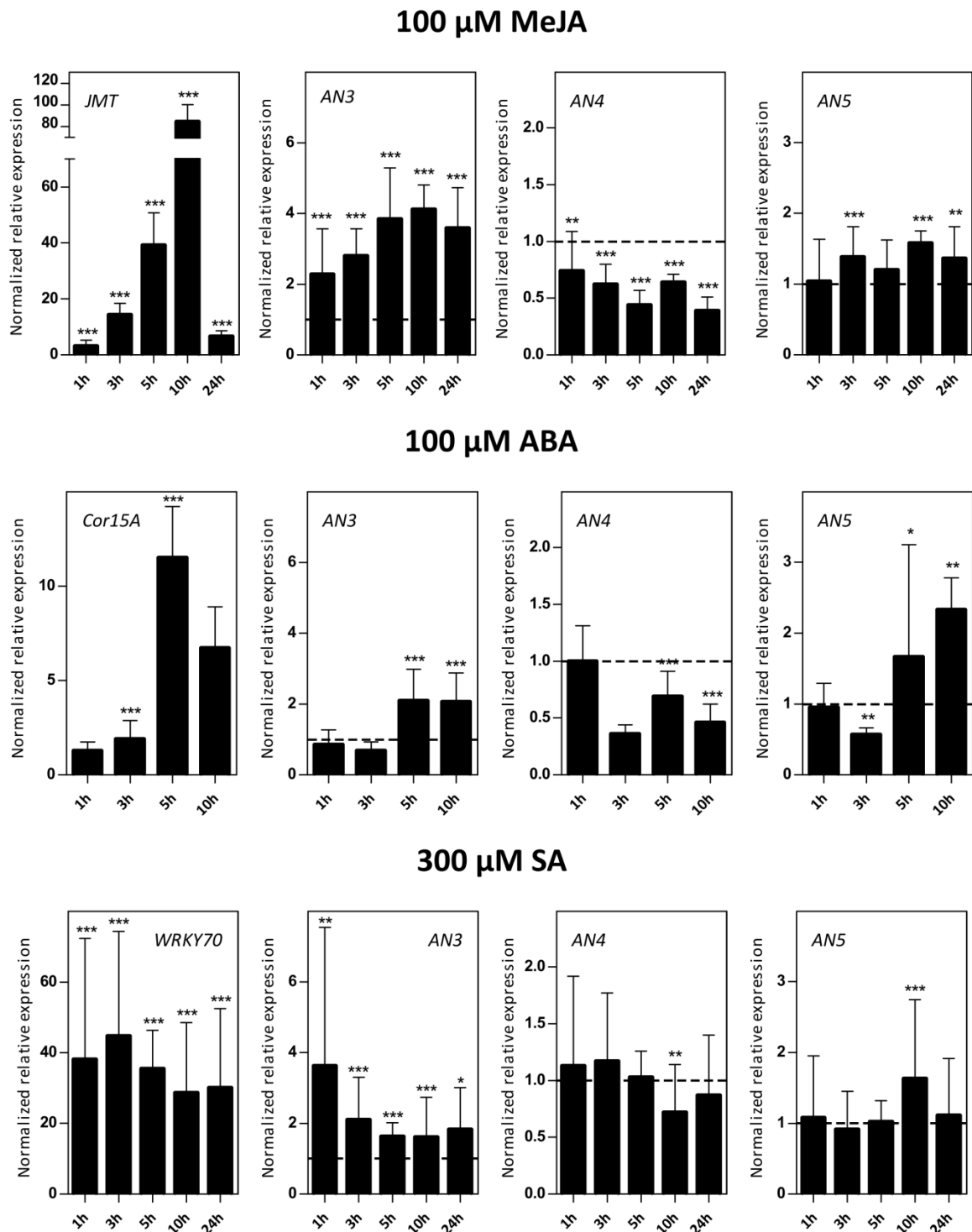


Figure 3.9 Normalized relative expression of the three *ArathNictaba* genes after MeJA, ABA and SA hormone treatment. The normalized transcript levels (to three reference genes: *PP2A*, *TIP41* and *UBC9*) are the result of two or four independent biological replicates (N = 2 for ABA, N = 4 for MeJA and SA) with similar results. They are presented relatively to the *ArathNictaba* expression level determined in the mock treated plantlets. Bars represent the mean \pm SE normalized relative expression and asterisks indicate statistically significant differences to the expression level of *ArathNictaba* in mock treated plantlets (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; REST analysis). The normalized relative expression levels for the positive control genes for each stress are presented in the left panels.

3.4.6.2 The expression of the *ArathNictabas* showed dissimilar patterns after abiotic stress treatments

Sixteen-day-old *Arabidopsis* seedlings were subjected to salinity (150 mM NaCl) and heat stress. Mock/abiotic stress treated samples were collected 1, 3, 5, 10 and 24 hours after treatment and transcript levels for *AN3*, *AN4* and *AN5* were determined by qPCR analysis (Figure 3.10). Control genes known to be responsive to salt (*RD29A*) and heat (*Hsp70b*) stress are significantly upregulated indicating that the plants sensed the different abiotic stress treatments. The expression level of *AN3* is not affected by salt stress but is 4 - 6 times significantly upregulated by heat stress after 3, 5, 10 and 24 hours. Overall, the expression of *AN4* is two times downregulated after salt as well as heat stress. The expression of *AN5* is only slightly influenced by salt stress and showed a threefold significant downregulation after 10 and 24 hours of heat stress.

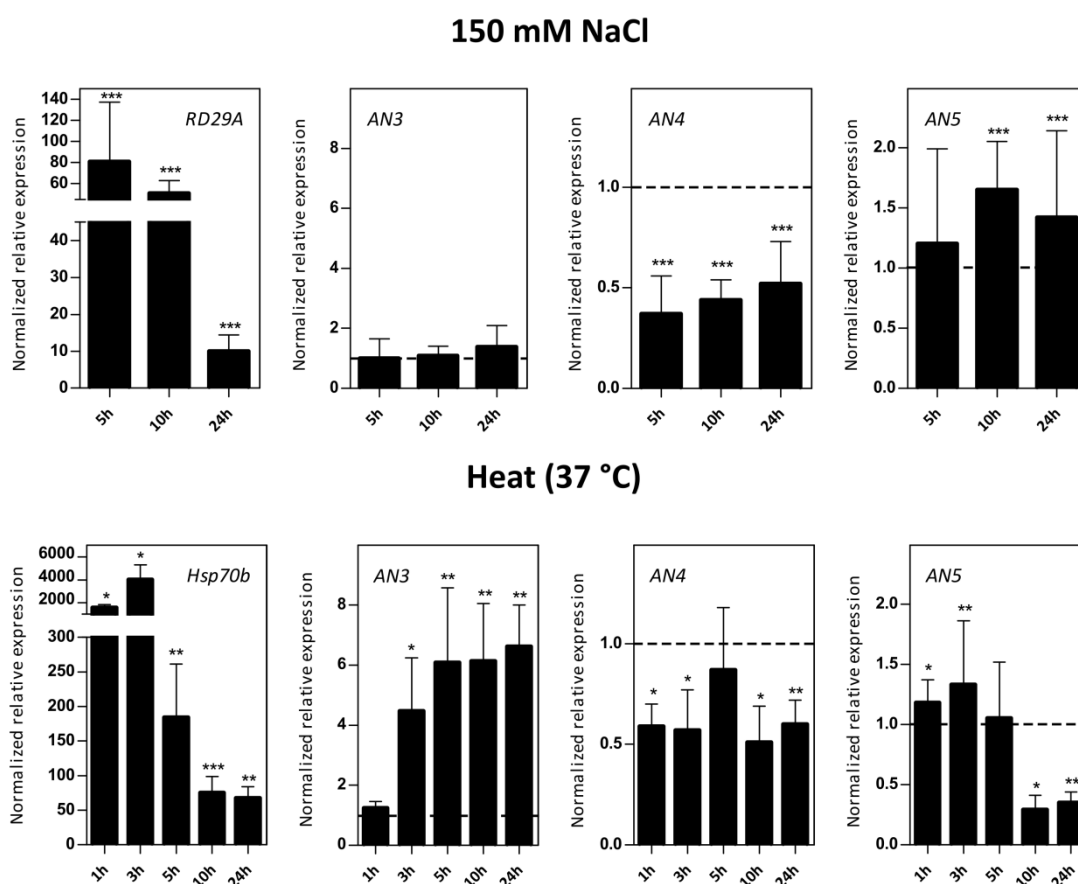


Figure 3.10 Normalized relative expression for the three *ArathNictaba* genes after salt and heat stress. The normalized transcript levels (to three reference genes: *PP2A*, *TIP41* and *UBC9*) are the result of two or four independent biological replicates ($N = 2$ for heat stress, $N = 4$ for salt stress) with similar results. They are presented relatively to the *ArathNictaba* expression level determined in the mock treated plantlets. Bars represent the mean \pm SE normalized relative expression and asterisks indicate statistically significant differences to the expression level of *ArathNictaba* in mock treated plantlets (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; REST analysis). The normalized relative expression levels of the positive control genes for each stress are presented in the left panels.

3.4.6.3 Expression of the *ArathNictabas* after different biotic stresses

Five-week-old *Arabidopsis* plants were subjected to *P. syringae*, *B. cinerea* infection and *M. persicae* infestation. Mock/biotic stress treated samples were collected 0 - 7 days post infection/infestation (dpi) and transcript levels of *AN3*, *AN4* and *AN5* were determined by qPCR analysis (Figure 3.11). This analysis revealed an early 2 - 2.5 fold upregulation (1 - 3 dpi) of the expression of *AN3* after *P. syringae* infection and a late twofold upregulation (5 - 7 dpi) of the expression of *AN4*. The expression of *AN5* is not changed by *P. syringae* infection. Fungal infection with *B. cinerea* affected *ArathNictaba* expression levels very weakly. Only a small significant downregulation 2 and 3 dpi for *AN5* was observed whereas infestation with *M. persicae* revealed an almost twofold upregulation after 3 days. The expression of *AN4* is not affected by infestation of the plants with *M. persicae* and the expression of *AN3* shows a twofold downregulation after 3 days. Control genes known to be responsive to *P. syringae* (*PR1*), *B. cinerea* (*PDF1.2*) and *M. persicae* (*PR1*) are significantly upregulated, indicating that the plants sensed the different biotic stress treatments.

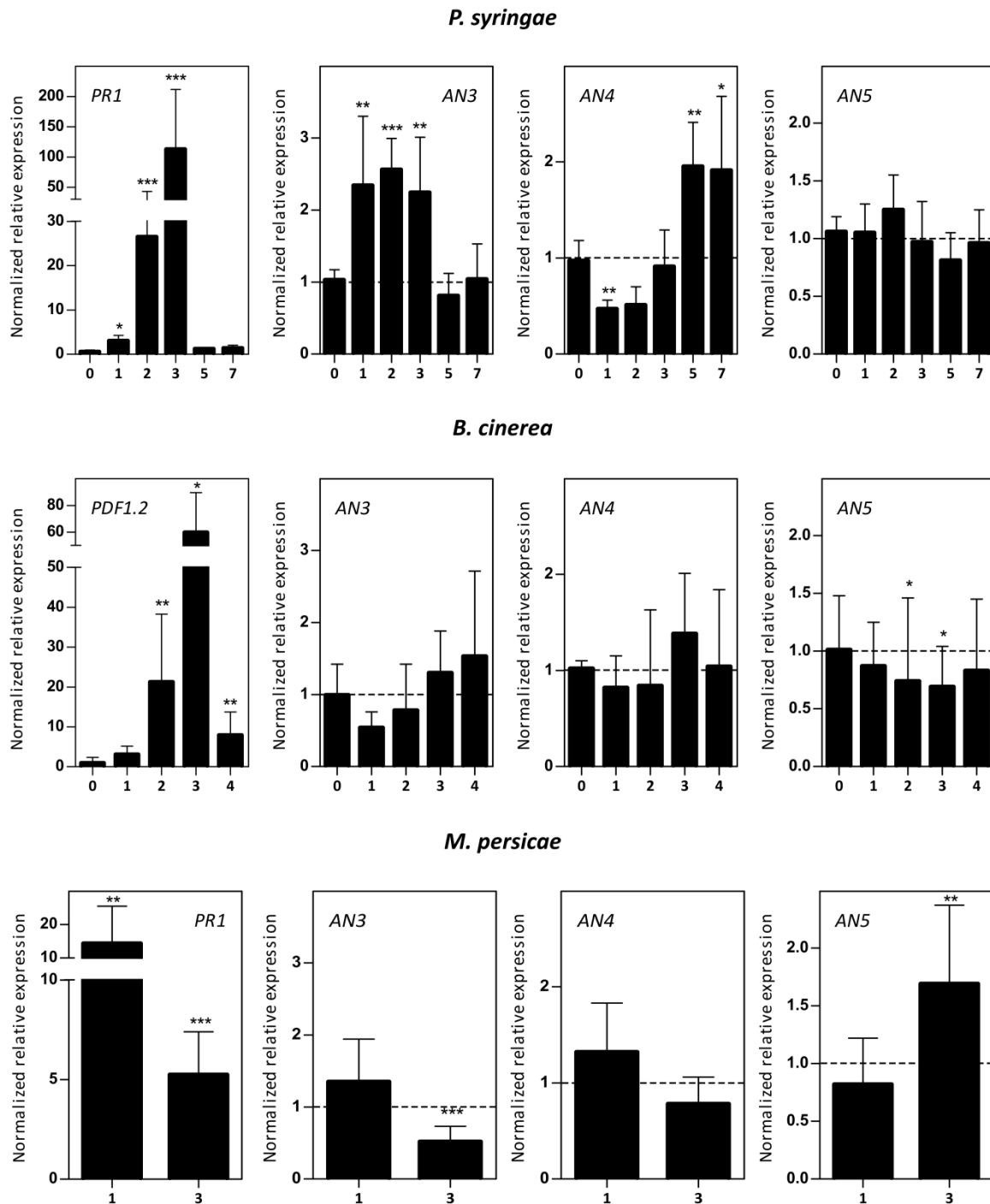


Figure 3.11 Normalized relative expression for the three *ArathNictaba* genes after *P. syringae* and *B. cinerea* infection and *M. persicae* infestation. The normalized transcript levels (to three reference genes: *PP2A*, *TIP41* and *UBC9*) are the result of two, three or four independent biological replicates (N = 2 for *B. cinerea*, N = 3 for *P. syringae*, N = 4 for *M. persicae*) with similar results. They are presented relatively to the *ArathNictaba* expression level determined in the mock treated plants. Bars represent the mean \pm SE normalized relative expression and asterisks indicate statistically significant differences to the expression level of *ArathNictaba* in mock treated plants (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; REST analysis). Numbers on the x axis represent the number of days after infection/infestation. The normalized relative expression levels of the positive control genes for each stress are presented in the left panels.

3.5 Discussion

The sequence alignment of Nictaba from tobacco and the Nictaba domains of the ArathNictaba sequences showed that the tryptophan residues important for carbohydrate binding in Nictaba are conserved in all ArathNictabas. Based on this observation, no conclusions can be drawn about the sugar specificity of the ArathNictabas. Stefanowicz *et al.* (2012) determined the sugar specificity of an F-box Nictaba (AT2G02360) protein from *A. thaliana*. The F-box Nictaba sequence contains the tryptophan residues important for sugar binding but glycan array analyses yielded results that were very different compared to the GlcNAc binding Nictaba. F-box Nictaba was shown to specifically bind to *N*-acetylglucosamine, Lewis A, Lewis X, Lewis Y and blood type B motifs (Stefanowicz *et al.*, 2012). Unfortunately, we did not succeed in determining the sugar specificity with a glycan array assay due to the low amount of recombinant ArathNictaba protein that could be purified (Chapter 5).

The localization pattern for all ArathNictabas was determined using EGFP fusion proteins. The results obtained by transient transformation of *N. benthamiana* were confirmed by stable transformation of *A. thaliana* (Figure 3.3, Figure 3.4). AN4 and AN5 showed a nucleocytoplasmic localization whereas AN3 only resided in the cytoplasm. The nucleocytoplasmic localization is similar to the localization pattern of Nictaba (Chen *et al.*, 2002; Lannoo *et al.*, 2006; Delporte *et al.*, 2014) and the localization pattern of two soybean Nictaba-like lectins (Van Holle *et al.*, 2016). The expression of EGFP fusion constructs in protein extracts from leaf tissue was analysed using sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) and Western blot analysis. In case of the transient transformation of *N. benthamiana*, no protein was detected with the anti-EGFP antibody, probably due to the low expression after transient transformation. In case of the stable transformation of *A. thaliana*, free EGFP was detected in some of the transgenic lines but no detectable levels of the intact fusion protein were observed with the anti-EGFP antibody. Possibly the extraction procedure was not suitable for the EGFP fusion proteins. The cytoplasmic localization of AN3-EGFP argues against the degradation of the fusion protein since free EGFP would localize in the nucleus.

No classical NLS is found in the ArathNictaba sequences suggesting that alternative mechanisms have to be used to enter the nucleus. In principle, all ArathNictaba EGFP fusion constructs are too big (> 40 kDa) to passively diffuse through nuclear pore complexes (Lange *et al.*, 2007). Apparently AN3-EGFP is not transported to the nucleus. Unfortunately, no fluorescence was ever detected for EGFP-AN3 which can have different possible causes going from transcriptional silencing to problems with protein folding.

Interestingly, the localization pattern of AN4 showed different results for N- and C-terminally tagged EGFP constructs in the transient transformation of *N. benthamiana*. EGFP-AN4 was clearly localized to the cytoplasm and the nucleus whereas AN4-EGFP was localized to the cytoplasm and only low fluorescence was observed in the nucleus. Apparently the efficiency

of nuclear import is much lower in the case of the C-terminal EGFP fusion construct of AN4. Differences in the localization pattern were also reported for one of the soybean Nictaba-like lectins (NLL3) (Van Holle, 2016). Strikingly in the case of AN4, this difference is not observed in the stable transformation of *A. thaliana* with the same EGFP constructs.

The SUBA3 server predicts the localization of a protein but also provides experimentally proven data of localization (Hooper *et al.*, 2014). Only for AN5, experiments with EGFP fusion constructs are mentioned and resulted in a nucleocytoplasmic localization (Cayla *et al.*, 2015). For AN3 and AN4, localization into the cytoplasm is predicted which does not totally agree with our experimental data.

Using qRT-PCR, the expression level of the *ArathNictabas* was investigated in different tissues from *A. thaliana* during development. The three *ArathNictaba* genes are expressed in every tissue during all developmental stages tested which is similar to the expression of *F-box Nictaba* and three soybean *Nictaba*-like lectins (Stefanowicz *et al.*, 2016; Van Holle, 2016; Van Holle *et al.*, 2016). However, these results are different from the expression data of *Nictaba* from tobacco which is not expressed under normal growth conditions (Chen *et al.*, 2002) suggesting that the function of the Nictaba homologs is more complex. The expression level of AN3 is higher in the stems and flowers compared to the other tissues which is comparable with the microarray data from eFP browser en Genevestigator (Winter *et al.*, 2007; Hruz *et al.*, 2008). Expression of AN3 was also determined in the seeds and pollen but these tissues were not sampled in our experiment. Expression of AN3 in the stems and the flowers is also in agreement with the AN3 ESTs identified in the green siliques and aboveground organs of two to six-week-old plants (Asamizu *et al.*, 2000). The expression of AN4 is higher in the roots and lower in the flowers compared to the other tissues according to the qPCR analysis. The expression in the roots is in agreement with the available microarray data which also show expression of AN4 in the stems, leaves and pollen but from the qPCR analysis, no significant differences for these tissues were demonstrated. Not coincidentally, the promoter analysis revealed the presence of some root meristem specific elements in the promoter of AN4. Moreover also AN4 ESTs confirm the expression of AN4 in the roots of *A. thaliana* (Asamizu *et al.*, 2000). The expression level of AN5 is higher in the rosette- and cauline leaves and slightly lower in the flowers compared to the other tissues. The expression of AN5 in the rosette- and cauline leaves is in agreement with the microarray data and the shoot specific elements present in the promoter sequence of AN5. Furthermore, ESTs for AN5 were identified in aboveground organs of two- to six-week-old plants (Asamizu *et al.*, 2000) and in vegetative tissues from adult plants. The qRT-PCR analysis revealed that the expression level of AN5 is higher than the expression of AN3 and AN4 which is in agreement with the microarray data (Supplementary figure 3). Furthermore qRT-PCR was performed with non-diluted cDNA for all executed qRT-PCRs for all genes. The raw Cq values with diluted cDNA for AN3 and AN4 were too high when the plants were not stressed meaning these *ArathNictabas* are expressed at very low levels.

In silico analyses suggested that the expression of all *ArathNictabas* is regulated by stress conditions. Data from eFP browser and Genevestigator revealed a highly fluctuating expression of *AN3* after different hormones, abiotic and biotic stress treatments (Winter *et al.*, 2007; Hruz *et al.*, 2008). Our qRT-PCR data revealed an overall upregulation after MeJA, ABA and SA treatment which is in agreement with the microarray data except for SA. Salt stress and *B. cinerea* infection had no influence on the expression level of *AN3*. In agreement with the microarray data, heat stress and *P. syringae* infection revealed an upregulated expression of *AN3*. *M. persicae* infestation resulted in downregulation of the expression of *AN3* while in the other tested stress conditions the expression of *AN3* is either upregulated or not influenced. The expression of *AN4* shows an overall downregulation after hormone and abiotic stress treatment which is not in agreement with the microarray data except for MeJA. Only *P. syringae* infection gave an upregulated expression of *AN4* which is again in contrast with the downregulated expression seen in the microarray data. The expression level of *AN5* is only slightly influenced by different hormones, abiotic and biotic stress treatments. The significant up- or downregulations are always less than twofold which is in agreement with the overall yellow coloured heat map for this *ArathNictaba* (Figure 3.6). Probably some of the discrepancies between the different datasets can be explained by differences in the experimental setup such as different time points, tissues and way of treatment.

The 2 - 2.5 fold upregulated expression of *AN3* and *AN4* after *P. syringae* infection is comparable with the expression of *F-box Nictaba* after *P. syringae* infection (Stefanowicz *et al.*, 2016). Next to that, *F-box Nictaba* also showed an upregulated expression after SA and heat stress treatment but only the expression of *AN3* is upregulated after these stresses. The expression of *AN4* is not influenced by SA treatment and downregulated by heat stress. Also, the expression pattern of *AN3* after salt stress is similar to the expression of *F-box Nictaba* after salt stress (Stefanowicz, 2015). The expression of both genes is not influenced by salt stress which is similar to the expression of *Nictaba* (Lannoo *et al.*, 2007a) but different from the expression of one of the soybean *Nictaba*-like lectins (*NLL1*) (Van Holle *et al.*, 2016). This comparison suggests that *AN3*, but not *AN4* might be involved in similar plant defence pathways as *F-box Nictaba*. In addition, *AN3* seems to play a role in MeJA and ABA stress. The expression of *AN3* is upregulated after treatments with these hormones while the expression of *F-box Nictaba* is not influenced by these hormones (Stefanowicz, 2015).

The expression level of *AN5* is upregulated three days after *M. persicae* infestation. Furthermore Beneteau *et al.* (2010) showed that recombinant *AN5* at mid-range concentrations, affects weight gain in *M. persicae* nymphs. The role of *AN5* in plant defence to *M. persicae*, particularly at the phloem-feeding stage, is further confirmed by *M. persicae* infestation experiments with transgenic lines overexpressing *AN5* (Zhang *et al.*, 2011).

Promoter analysis revealed a large amount of different cis-regulatory elements in the *ArathNictaba* promoter sequences. Although this analysis can give indications on the

regulation of gene expression, these data need to be confirmed experimentally. The promoter of *AN3* is the only promoter that contains a heat responsive element and microarray data suggest an upregulated expression of *AN3* after heat stress (Figure 3.5, Figure 3.6). In our data, *AN3* is indeed the only *ArathNictaba* for which the expression is upregulated after heat stress (Figure 3.10). In contrast, *ArathNictaba* promoters for *AN4* and *AN5* have MeJA responsive elements in their promoter sequence contrary to the promoter of *AN3* while microarray data revealed an overall upregulation of the expression of *AN3* after MeJA treatment (Figure 3.5, Figure 3.6). In agreement with the microarray data, this is the only *ArathNictaba* for which the expression is strongly upregulated after MeJA treatment (Figure 3.9). Next, *AN3* is the only *Nictaba* homolog for which the expression is upregulated by MeJA treatment, the expression of *F-box Nictaba* and *Nictaba*-like soybean lectin genes is not influenced by MeJA stress (Stefanowicz, 2015; Van Holle, 2016). The downregulated expression of *AN4* after 3 hours MeJA treatment seen in the microarray data can also be confirmed by our expression data which show an overall downregulation of *AN4* after MeJA treatment (Figure 3.6, Figure 3.9). All *ArathNictaba* promoters contain SA responsive elements whereas microarray data revealed only a small downregulation of *AN3* after SA treatment (Figure 3.5, Figure 3.6). This downregulation is in contrast with our data which revealed an overall upregulation of the expression of *AN3* after SA stress (Figure 3.9).

As illustrated in this chapter, the different *ArathNictabas* show different expression patterns under normal growth conditions and after different stress treatments. The differential expression of these *ArathNictabas* suggests that different homologs can have a complementary role in the plant response to stress conditions. The comparison of the *ArathNictaba* expression patterns with the expression patterns previously reported for *F-box Nictaba* from *A. thaliana* and the soybean *Nictaba*-like lectins indicates that all *Nictaba* homologs are involved in different stress related pathways. Considering the role of the *Nictaba* domain in the plant stress response, the performance of transgenic lines overexpressing the *ArathNictaba* genes in stress conditions will be investigated in the next chapter.

Chapter 4

**Functional study of Nictaba homologs in
*Arabidopsis thaliana***

4.1 Abstract

Plant lectins constitute an important part of the different plant immune receptors located either at the cell-surface or in the cytoplasmic compartment. Nictaba was shown to play a role in plant stress responses and the expression of *ArathNictabas* is stress-inducible indicating these Nictaba homologs also play a role in plant stress responses. In this chapter *ArathNictaba* overexpression lines were generated and subjected to salt stress and *P. syringae* infection in order to elucidate whether the *ArathNictaba* genes give rise to more tolerant plants in these stress situations. Germination experiments with *ArathNictaba* overexpression lines under salt stress conditions revealed no better tolerance of the transgenic plants towards mild and high salt stress compared to WT plants. Judging from the *P. syringae* infection experiments with the *ArathNictaba* overexpression lines, several transgenic lines for AN4 and AN5 showed a significantly lower level of leaf damage, percentage of cell death and Pseudomonas biomass. It can be concluded that two of the three Nictaba homologs possibly play a role in the defence of the plant against *P. syringae*.

4.2 Introduction

Plants are constantly exposed to different abiotic and biotic stresses. The innate immune system protects and defends plants by using different immune receptors located either at the cell-surface or in the cytoplasmic compartment. Plant lectins constitute an important part of these receptors. The cell-surface immune receptors with a lectin domain are part of the GNA, legume and LysM lectin families. The intracellular immune receptors with a lectin domain are nucleocytoplasmic lectins (Lannoo and Van Damme, 2014; Eggermont *et al.*, 2017 (Chapter 2)). At present, members of the amaranthin, EUL, GNA, JRL, Nictaba and ricin B lectin families have been studied and shown to play a role in plant stress responses (Van Damme *et al.*, 2001, 2004; Wu *et al.*, 2006; Vandenborre *et al.*, 2010; Xin *et al.*, 2011; Al Atalah *et al.*, 2014; Song *et al.*, 2014; Van Hove *et al.*, 2015).

The Nictaba lectin family was named after the *Nicotiana tabacum* agglutinin, abbreviated as Nictaba, originally discovered in tobacco leaves (Chen *et al.*, 2002). The level of Nictaba expression was investigated after several hormone, abiotic and biotic stress treatments. Only jasmonate, MeJA, 12-oxo-phytodienoic acid (jasmonate precursor), chewing caterpillars and cell-content-feeding spider mites enhanced Nictaba synthesis and expression in tobacco leaves (Lannoo *et al.*, 2007a; Vandenborre *et al.*, 2009). Furthermore, feeding experiments with tobacco lines in which Nictaba was overexpressed or silenced, demonstrated the insecticidal activity of Nictaba. It was suggested that the interaction of Nictaba with glycoconjugates present in the digestive tract of the insects caused the entomotoxic activity (Vandenborre *et al.*, 2010, 2011). Upon insect herbivory, the expression of Nictaba within the plant cell is enhanced and the protein is localized to the cytoplasm and the nucleus. It is believed that Nictaba binding to O-GlcNAc modified core histones results in chromatin remodelling and as such in enhanced transcription of defence related genes (Schoupe *et*

al., 2011; Delporte *et al.*, 2014). Research on one of the *ArathNictabas* also showed its critical role in the *Arabidopsis* defence system. Beneteau *et al.* (2010) showed that recombinant AN5 at mid-range concentrations, affects weight gain in *M. persicae* nymphs and Zhang *et al.* (2011) revealed a repression in phloem-feeding of *M. persicae* on AN5 overexpression lines. Furthermore, growth of various strains of filamentous fungi was inhibited by recombinant AN5 protein (Lee *et al.*, 2014).

In order to refine the function of an unknown gene/protein, BLAST analyses can reveal sequence similarity with other proteins of which the function is known. Similarly, protein domain identification will give indications about possible functions. Furthermore, ESTs can give additional information and support the information obtained with the BLAST analyses and protein domain searches (Swarbreck *et al.*, 2008). When and where a gene is expressed in the cell or in the whole organism can be determined experimentally via different techniques and can give important clues about the function of a gene. The most direct way to find out the function of a gene, is studying mutant and overexpression lines (Alberts *et al.*, 2002). Loss-of-function mutants are usually created by random mutations by chemicals or by a T-DNA insertion. However, characterization of many single-gene mutants is challenging in plants because of genetic redundancy. Gain-of-function mutants can be created by random insertions of transcriptional enhancers or expression of transgenes under the control of a strong constitutive promoter. In contrast to single-mutants, overexpression of a gene can characterize functionally redundant genes (Çiftçi, 2012).

As shown in chapter 3, the expression of the *ArathNictabas* is stress-inducible. In order to refine our understanding of the role of these *Nictaba* homologs in the plant, *ArathNictaba* overexpression lines were generated and tested under normal growth conditions, upon salt treatment and after *P. syringae* infections. These experiments allowed us to investigate if overexpression of the *ArathNictaba* genes leads to a better tolerance towards salt stress and reduced susceptibility towards *Pseudomonas* infection compared to WT plants.

4.3 Materials and methods

4.3.1 Plant material and growth conditions

WT *A. thaliana* seeds, ecotype Col-0, were purchased from Lehle Seeds (Round Rock, Texas, USA). *Arabidopsis* seeds were grown *in vitro*, in pots containing commercial soil or individually grown in Jiffy's. After a 3 days stratification period at 4 °C in the dark, plants were grown at 21 °C in a controlled growth chamber with a 16/8 h photoperiod. All plant materials and growth conditions were previously described in chapter 3, section 3.3.1.

4.3.2 Construction of *ArathNictaba* overexpression constructs

The entry clones containing the *AN3*, *AN4* and *AN5* coding sequences with stop codon (Chapter 3, section 3.3.7) were used in an LR reaction to generate the overexpression constructs. The destination vector pK7WG2,0 (Karimi *et al.*, 2002) containing the 35S promoter was combined with the entry clones and the Gateway LR Clonase II to get the desired expression clones (35S::*AN3*, 35S::*AN4* and 35S::*AN5*). These expression clones were heat shock transformed in TOP10 *E. coli* cells and transformants were selected on LB agar plates containing 75 µg/mL spectinomycin. Transformants were screened with colony PCR using a forward primer in the 35S promoter and a reverse primer in the 35S terminator sequence (Supplementary table 7). Electrocompetent *A. tumefaciens* GV3101 were transformed with these expression clones (300 ng) via electroporation with the following parameters: 2.0 kV, 25 µF and 200 Ω. Immediately after the pulse, YEB medium was added to the cells and they were grown on a shaker (200 rpm) at 28 °C for two hours. Transformants were selected on YEB agar plates with 75 µg/mL spectinomycin and screened with colony PCR using primers located in the 35S promoter and terminator sequences (Supplementary table 7).

4.3.3 Stable transformation of *A. thaliana* plants

Stably transformed *Arabidopsis* plants were created using the floral dip transformation method (Clough and Bent, 1998). Transformed seeds were selected using the fast selection protocol according to Harrison *et al.* (2006). Selection was done until the transformed plants were homozygous for the T-DNA integration (T3 and T4). Integration of the T-DNA was checked by PCR on gDNA using primers in the *kanamycin* resistance gene (Supplementary table 7) using the following PCR program: 10' 94 °C – 45 x (30" 94 °C - 30" 48 °C - 1' 72 °C) - 5' 72 °C. Simultaneously, the quality of the gDNA was checked with actin primers (Supplementary table 7) using the same PCR program. Overexpression levels of the *ArathNictaba* genes in 15-day-old seedlings were quantified by qRT-PCR. Three independent homozygous single insertion lines of each construct (35S::*AN3*, 35S::*AN4* and 35S::*AN5*) were selected and used in all experiments.

4.3.4 RNA extraction, cDNA synthesis, RT-PCR and qRT-PCR analysis

Arabidopsis seedlings for the different overexpression lines were ground in liquid nitrogen for total RNA isolation, followed by cDNA synthesis and RT-PCR analysis as previously described (Chapter 3, section 3.3.6). qRT-PCR analyses were performed using the 96-well CFX Connect™ Real-Time PCR Detection System (Bio-Rad) as previously described (Chapter 3, section 3.3.11).

4.3.5 Checking T-DNA insertion lines for *AN4* and *AN5*

Two mutant lines were available from the Nottingham *Arabidopsis* Stock Centre (NASC), one with a T-DNA insertion in the promoter of *AN4*, the other one with a T-DNA insertion in the

promoter of *AN5*. T-DNA insertion lines for *AN3* were not available. The mutant line for *AN4* is a SALK line created by the Ecker lab at the Salk Institute (<http://signal.salk.edu/cgi-bin/tdnaexpress>). The mutant line for *AN5* is a SAIL line created by Syngenta (North Carolina, USA). Plants grown in soil were tested for the T-DNA insertion by PCR on total gDNA. Total gDNA was extracted using the slightly adapted Edwards protocol (Edwards *et al.*, 1991). After extraction, gDNA concentrations were measured with the Nanodrop 2000 spectrophotometer (Thermo Scientific) and diluted to 100 ng/μL. For each T-DNA insertion line, three primers were used, including a left border primer on the T-DNA insertion as well as two primers surrounding the T-DNA insertion (left primer (LP) and right primer (RP)) (Supplementary table 7). Two PCR reactions were performed with the following combinations of primers for each T-DNA insertion line: LP + RP and left border primer (LB) + RP. These PCR reactions were also performed on gDNA extracted from WT *A. thaliana* plants. Simultaneously, the quality of the gDNA was checked by PCR using *ACT2* primers (Supplementary table 7). PCR reactions included 200 ng gDNA, 2 μL 10 mM dNTPs (Thermo Fisher Scientific), 1 μL of each primer (5 μM, Life Technologies), 2.5 μL 10 x Extra buffer (VWR), 0.125 *Taq* DNA polymerase (VWR) and water up to the volume of 25 μL. The PCR conditions are as follows: 2' 94 °C – 30 x (15" 94 °C - 30" 47 or 52 °C - 1'30" 72 °C) - 5' 72 °C, with the annealing temperature 47 or 52 °C depending on the primers used.

4.3.6 Germination assays and salt stress tolerance

Arabidopsis seeds of WT plants and three independent homozygous transgenic lines with overexpression of *AN4* and *AN5* (35S::*AN4* and 35S::*AN5*) were grown on 1/2 MS medium containing 0, 50 or 150 mM NaCl. Fifty seeds per transgenic line per treatment were sown. After stratification (3 days, 4 °C in the dark) the plates were placed in a controlled plant growth room at 21 °C and a 16/8 h light/dark photoperiod. Germination was followed for seven days and seeds were counted as germinated when the radicle emerged through the seed coat. Two independent biological replicates were performed. Results were statistically analysed with SPSS Statistics 22 (IBM) using the Pearson chi-square test for binomial distributed data and data were considered significant for $p \leq 0.05$.

4.3.7 Biotic stress susceptibility experiment

The *Pseudomonas syringae* pv. *tomato* DC3000 strain was supplied by the Phytopathology lab of Prof. dr. M. Höfte (Ghent University, Belgium). Infection assays were performed according to Pieterse *et al.* (1996), Audenaert *et al.* (2002) and Katagiri *et al.* (2002), with some minor modifications as described in chapter 3, section 3.3.5. Four-week-old *Arabidopsis* WT plants and three independent homozygous transgenic lines for each construct (35S::*AN3*, 35S::*AN4* and 35S::*AN5*) were inoculated with the infection or mock solutions. Rosette leaves of three individual randomly chosen plants were sampled at different time points post infection. Two independent biological replicates were performed and in each replicate, samples were used to measure leaf damage (section 4.3.7.1), visualize cell death (section 4.3.7.2) and determine *Pseudomonas* biomass (section 4.3.7.3). Samples

to determine *Pseudomonas* biomass were stored at -80 °C prior to DNA extraction and qRT-PCR analysis.

4.3.7.1 Quantification of leaf damage

To measure leaf damage, six leaves per line per time point were scanned with a flatbed scanner at the highest resolution. The percentage of leaf damage (relative to the total leaf area) was determined in the Image Analysis Software for Plant Disease Quantification Assess 2.0 (APS, St. Paul, USA) using a self-written macro adjusted to our sampled leaves. The data were tested for normal distribution with the Shapiro-Wilkinson test. The Mann-Whitney U test was used for not normal distributed data, supplemented with a non-parametric equivalent of the Levene's test to check homogeneity of variances. The Bonferroni-Holm correction was used for multiple testing.

4.3.7.2 Visualisation and quantification of cell death

Using trypan blue staining (Sigma-Aldrich, Diegem, Belgium), plant cell death was visualized on infected and mock treated leaves from the *P. syringae* infection experiment. One leaf of three randomly chosen plants per line for each time point was submerged with the trypan blue solution (0.02 %) and boiled for two minutes. Then, the boiled leaves in trypan blue solution were incubated overnight at room temperature on a rotary shaker (50-100 rpm). Next day, the trypan blue solution was replaced by a chloral hydrate solution (100 g / 40 mL water) to destain the leaves. The destained leaves were placed on a microscopy slide in 50 % glycerol and pictures were taken with a Leica S8APO microscope (DFC400 camera) and Leica Plan APO 1.6 x objective.

The trypan blue staining was quantified by a scoring system. Each picture was scored by estimating the percentage of blue staining or cell death. Leaves without trypan blue staining (0 %) were assigned a score 1. Score 2 was assigned to leaves for which the percentage of cell death was 1 - 30 %. Leaves with 31 - 60 % of cell death were assigned a score 3. Score 4 was the highest score and was assigned to leaves for which the percentage of cell death was 61 - 100 %. Each transgenic line was statistically compared with the WT using a Mann-Whitney U test supplemented with a non-parametric equivalent of the Levene's test to check homogeneity of variances. The Bonferroni-Holm correction was used for multiple testing.

4.3.7.3 Determination of *P. syringae* biomass

To determine the *P. syringae* biomass, gDNA was first extracted from the infected and mock treated leaves using hexadecyltrimethylammonium bromide (CTAB). The CTAB buffer (2 % CTAB, 0.1 M Tris-HCl pH 7.5, 1.4 M NaCl, 20 mM EDTA) was added to 100 mg of plant material and this mixture was incubated for 90 minutes at 65 °C in a shaking heat block. Next, the extraction was followed by a chloroform:isoamylalcohol (24:1) precipitation. Finally, the gDNA was precipitated with 100 % isopropanol, washed with 76 % ethanol/0.2 M

NaOAc and 76 % ethanol/10 mM NH₄OAc and dissolved in water. Quantification of *P. syringae* biomass was performed with qRT-PCR using *oprF* primers targeting the outer membrane porin protein F gene of *P. syringae* (Brouwer *et al.*, 2003) (Supplementary table 5). *ACT2* and *PEX4* primers were used as reference genes from *A. thaliana* (Supplementary table 5). The REST-384 software was used to calculate the ratio of *P. syringae* gDNA to *A. thaliana* gDNA (Pfaffl *et al.*, 2002).

4.4 Results

4.4.1 Selection of T-DNA insertion and overexpression lines

4.4.1.1 Available AN4 and AN5 T-DNA insertion lines contain no T-DNA insertion

The SALK T-DNA insertion lines were generated by Dr. Joseph Ecker (Salk Institute, California, USA) via *A. tumefaciens* vacuum infiltration of *A. thaliana* ecotype Col-0 (Alonso *et al.*, 2003). The SAIL T-DNA insertion lines were generated by Syngenta (North Carolina, USA), also via *A. tumefaciens* vacuum infiltration of *A. thaliana* ecotype Col-0 (McElver *et al.*, 2001). Two available T-DNA insertion lines, one for AN4 (SALK_019483), one for AN5 (SAIL_835_C05), were obtained from NASC and the T-DNA insertion was analyzed. For both lines, the T-DNA insertion is situated in the promoter of the gene. PCR was performed on total gDNA using the LB on the T-DNA insertion as well as the LP and RP spanning the insertion site (Figure 4.1) (Supplementary table 7). As shown in Figure 4.2 in the lanes with the *ACT2* primers, the quality of the gDNA for each sample (WT, SALK and SAIL line) was approved, the expected band of 390 bp is present in all gDNA samples. By using LP and RP, the primers spanning the T-DNA insertion site, bands of 1069 bp and 1508 bp were amplified for SALK_019483 (AN4) and SAIL_835_C05 (AN5) respectively, but the same bands were amplified using LP and RP on the gDNA from WT plants. This result suggests that the T-DNA insertion lines contain no T-DNA insertion. This result is confirmed in the PCR with primer combination LB + RP as there is no PCR product amplified. These data indicate that the left border primer is not able to bind on the T-DNA insertion which confirms the T-DNA insertion is absent in both T-DNA insertion lines. As such these lines cannot be used in further experiments.

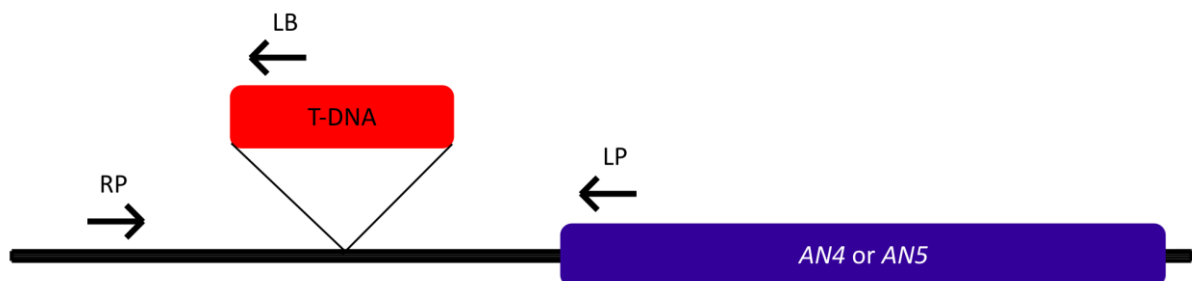


Figure 4.1 Positions of primers to analyse T-DNA insertion lines for AN4 and AN5. Left primer (LP), right primer (RP) and left border primer (LB).

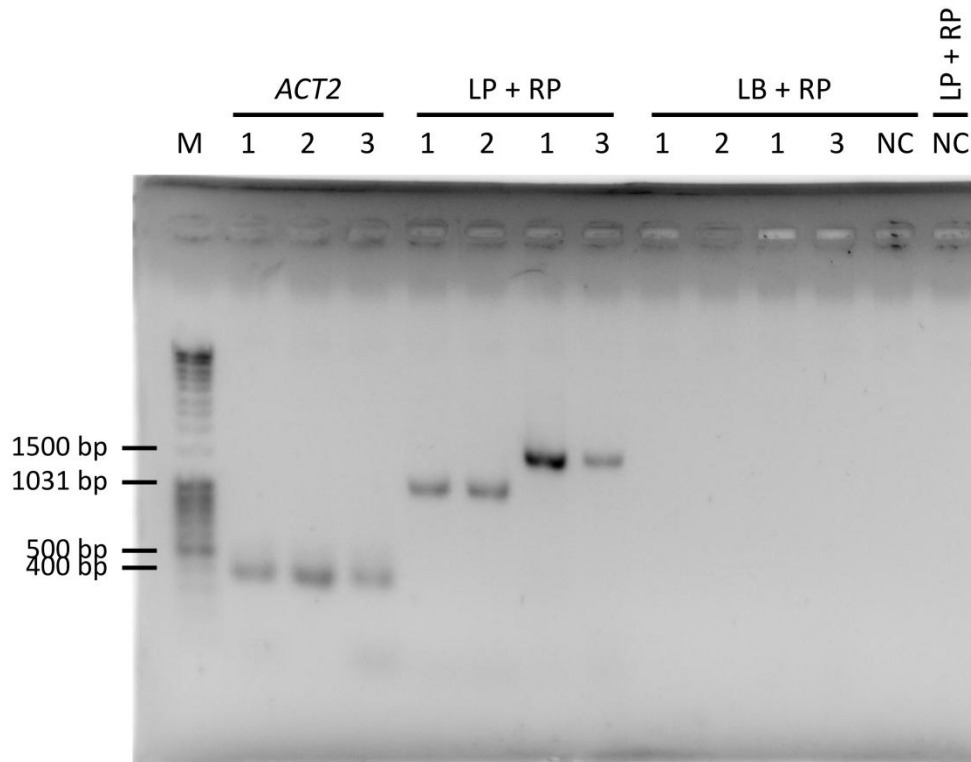


Figure 4.2 Analysis of T-DNA insertion lines for AN4 and AN5. PCR on gDNA extracted from WT plants (1), AN4 T-DNA insertion line SALK_019483 (2), AN5 T-DNA insertion line SAIL_835_C05 (3) compared to water, negative control (NC). Primer combinations used are indicated with left primer (LP), right primer (RP) and left border primer (LB). ACT2 refers to a forward and reverse primer on the ACT2 gene. M: DNA marker.

4.4.1.2 Overexpression lines for AN3, AN4 and AN5

A. thaliana ecotype Col-0 was transformed with the overexpression constructs 35S::*AN3*, 35S::*AN4* and 35S::*AN5* and transgenic lines were selected. The resulting homozygous *AN3*, *AN4* and *AN5* overexpression lines carrying a single copy of the T-DNA insertion were tested on transcript level with qRT-PCR using cDNA from 15-day-old *Arabidopsis* seedlings. Based on their expression level relative to the *ArathNictaba* expression in WT plants, three independent overexpression lines exhibiting different overexpression levels were selected (Figure 4.3). Taking into account the higher expression levels of *AN5* in WT plants compared to the expression levels of *AN3* and *AN4* (Chapter 3, section 3.4.5), it is not surprising to observe a lower overexpression level for the 35S::*AN5* transgenic lines (Figure 4.3). The expression level of *AN3* in the overexpression lines is more than 400 times higher than in WT plants for all three selected overexpression lines. In contrast with that, two of the three overexpression lines for *AN4* have an expression level of *AN4* which is only 50 and 300 times higher than in WT plants. All overexpression lines were also checked on protein level by SDS-PAGE and Western blot analysis with an anti-Nictaba antibody. The coomassie stained protein gels revealed no obvious differences in protein patterns between the protein extracts of WT plants and overexpression lines (data not shown).

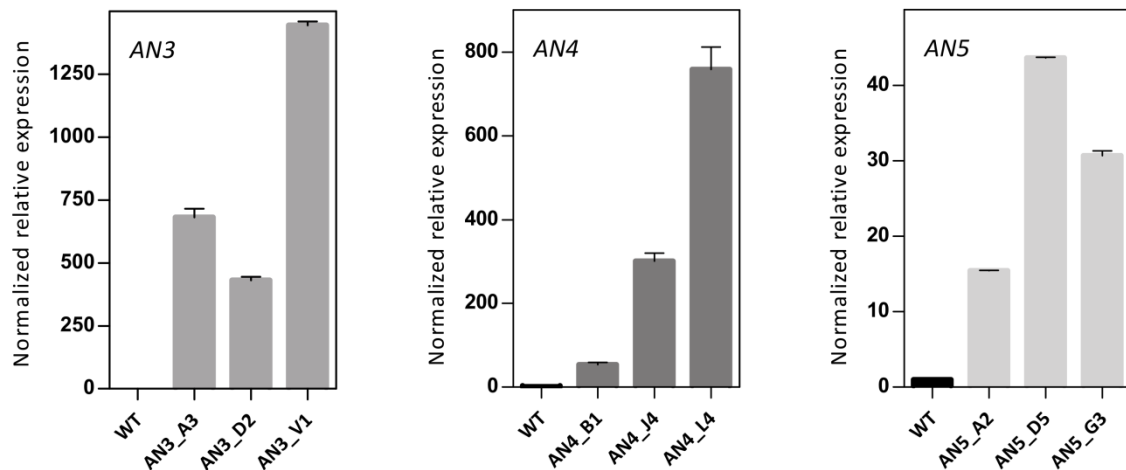


Figure 4.3 Expression analysis of *ArathNictaba* overexpression level in 2-week-old 35S::AN3, 35S::AN4 and 35S::AN5 transgenic lines. Normalized relative expression of AN3, AN4 and AN5 compared to their expression in WT plants (N = 1). Error bars represent standard deviations.

4.4.2 Germination assay of different transgenic lines compared to WT *A. thaliana* under normal growth and salt stress conditions

To check whether the *ArathNictaba* genes play a role in the tolerance of the plant against salt stress, the germination of the 35S::AN4 and 35S::AN5 transgenic lines was tested on 1/2 MS medium supplemented with different salt concentrations.

The germination of all overexpression lines except AN4_L4 on day seven after a stratification period of 3 days is significantly lower than the germination of the WT seeds on MS medium without salt (Figure 4.4). On MS medium supplemented with 50 mM NaCl, the germination of AN4_J4 and AN4_L4 is not significantly different from the germination of the WT seeds. Taking into account the significantly lower germination of AN4_J4 on MS medium without salt, this suggests AN4_J4 may have a better tolerance against mild salt stress (50 mM NaCl). On MS medium with 150 mM NaCl, the germination percentage of all overexpression lines is significantly lower than the germination percentage of the WT seeds. None of the overexpression lines is performing better on MS medium with 150 mM NaCl than the WT plants. As such it can be concluded that all the overexpression lines are more sensitive to high salt stress (150 mM NaCl).

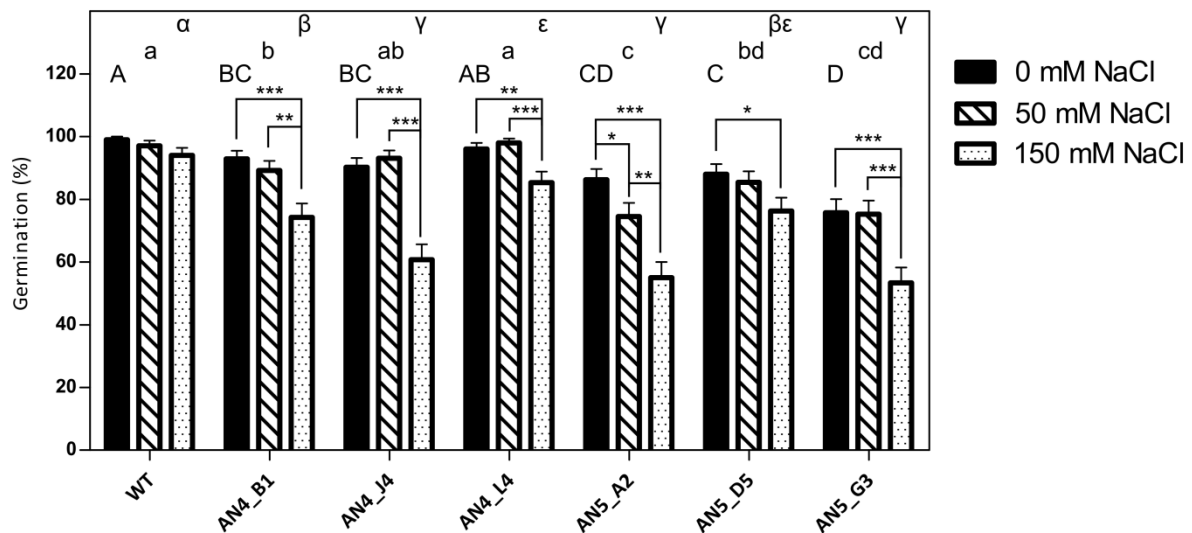


Figure 4.4 Percentage germination of WT and transgenic overexpression lines on 1/2 MS medium supplemented with different NaCl concentrations on day 7 after a stratification period of 3 days. Bars represent the mean \pm SE of two independent biological replicates ($N = 2$) with fifty seeds per line per replicate. Asterisks represent statistically significant differences within one line between the different salt concentrations (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; Pearson chi-square test). Different letters represent statistically significant differences within one salt concentration between the different lines (* $p \leq 0.05$, Pearson chi-square test).

4.4.3 *ArathNictaba* overexpression lines show less disease symptoms and bacterial growth after *P. syringae* infection

WT and transgenic 35S::*AN3*, 35S::*AN4* and 35S::*AN5* *A. thaliana* plants were infected with the virulent hemibiotrophic *P. syringae* to investigate the role of the *ArathNictaba* genes in the defence against this pathogen. Infection of *A. thaliana* plants with *P. syringae* results in yellow lesion areas on the rosette leaves of the plant. This leaf damage was measured daily on scanned leaves and the percentage of the lesion area relative to the total leaf area was calculated. First bacterial lesions started to appear at three dpi, but only at four dpi differences in leaf damage were observed for the overexpression lines compared to WT plants (Figure 4.5). Statistically significant differences in leaf damage compared to the leaf damage in WT plants are observed especially for lines AN4_B1, AN4_L4, AN5_D5 and AN5_G3. These four overexpression lines reveal a significantly lower percentage of leaf damage compared to WT plants suggesting they are more tolerant to *P. syringae* infection. The leaf damage in mock treated plants was also determined but was never higher than 6.5 % (data not shown). Furthermore, the leaf damage of the 35S::*AN5* lines is significantly ($p \leq 0.05$) negatively correlated with the expression level of the different overexpression lines (Pearson correlation, SPSS23). No correlation was observed between the leaf damage and the expression level of *AN3* and *AN4* in the different overexpression lines for *AN3* and *AN4*.

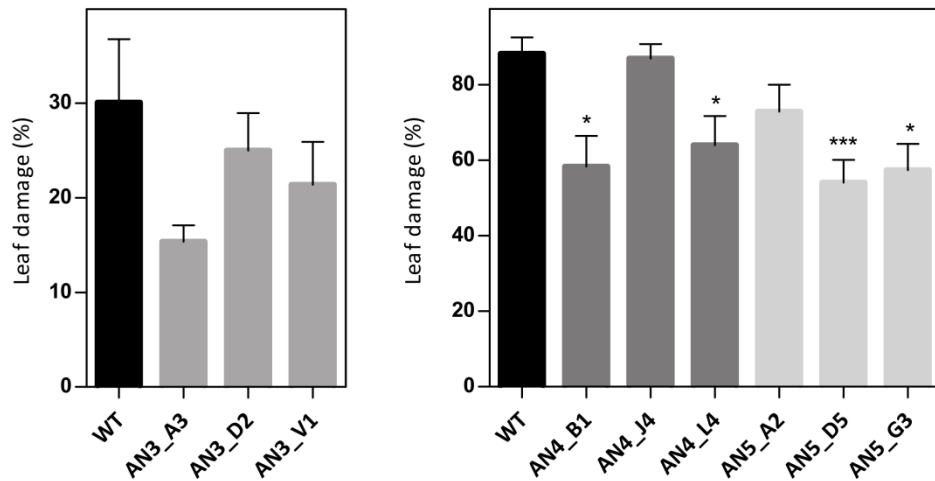


Figure 4.5 Percentage of leaf damage in WT and transgenic overexpression *A. thaliana* plants infected with *P. syringae* at 4 dpi. Percentage of leaf damage calculated with Assess 2.0 at 4 dpi as percent ratio of yellow lesion area relative to the total leaf area. Bars represent the mean \pm SE of two independent biological replicates with six individual leaves per line per replicate. Asterisks indicate statistically significant differences to the percentage of leaf damage in WT plants (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; Mann-Whitney U test).

To strengthen these results, bacterial growth on the *A. thaliana* WT and transgenic plants was compared in a qPCR analysis to quantify the relative *P. syringae* biomass. *P. syringae* and *A. thaliana* specific primers were used in this qPCR analysis on gDNA extracted from leaf material collected at three and four dpi. AN4_B1, AN5_D5 and AN5_G3, which had a significantly lower percentage of leaf damage compared with WT plants, also have a significantly lower *P. syringae* biomass (Figure 4.6). These data confirm that these overexpression lines are more tolerant than WT plants to infection with *P. syringae*. Even AN5_A2, which had less (though not significant) leaf damage compared to WT, revealed a significantly lower *P. syringae* biomass. AN4_L4 has a lower (but not significant) *P. syringae* biomass compared to WT plants. All AN3 overexpression lines show a higher *P. syringae* biomass compared to WT plants, which is in contrast with the lower (but not significant) leaf damage compared to WT plants. It cannot be concluded that AN3 overexpression lines are more tolerant to infection with *P. syringae*.

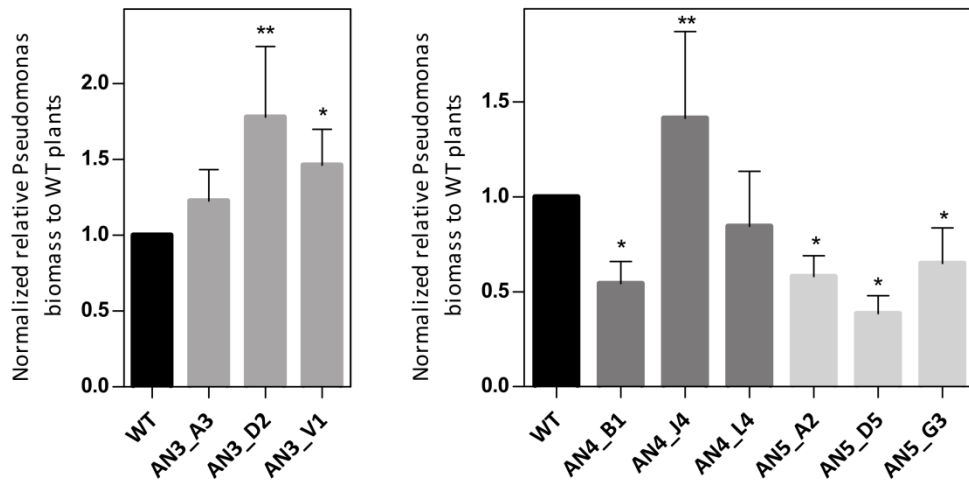


Figure 4.6 Normalized relative *P. syringae* biomass in the overexpression lines compared to WT plants at 4 dpi. Bars represent the mean \pm SE relative *P. syringae* biomass in overexpression lines to biomass in WT plants at 4 dpi from two independent biological replicates normalized with two *A. thaliana* reference genes (*ACT2* and *PEX4*) in REST-384. Asterisks indicate statistically significant differences to the biomass in WT plants (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; pair wise fixed reallocation randomization test REST-384).

A third analysis includes the visualization and quantification of cell death in the leaves using trypan blue staining (Figure 4.7, Figure 4.8). No significant differences were obtained for the percentage of cell death of 35S::*AN3* lines compared to WT plants, this result confirms the result obtained by leaf damage analysis but is also in contrast with the higher *P. syringae* biomass of the *AN3* overexpression lines compared to WT plants. *AN4_B1* showed a significantly lower leaf damage and *P. syringae* biomass compared to WT plants (Figure 4.5, Figure 4.6) which is in agreement with the difference (not significant) in percentages of cell death (Figure 4.7) concluding *AN4_B1* is more tolerant to *P. syringae* infection than WT plants. Surprisingly *AN4_J4* shows a significantly reduced amount of cell death compared to WT plants (Figure 4.7) as the percentage of leaf damage was almost equal to the WT plants and the *P. syringae* biomass was significantly higher (Figure 4.5, Figure 4.6). *AN4_L4* shows comparable results to WT plants in all analyses except for the leaf damage analysis which revealed a significantly lower leaf damage (Figure 4.5). As such it cannot be concluded that *AN4_J4* and *AN4_L4* are more tolerant to *P. syringae* infection. *P. syringae* biomass for *AN5_A2* is significantly lower than for the WT plants (Figure 4.6) which is in agreement with the lower trend obtained by leaf damage quantification and amount of cell death suggesting *AN5_A2* is more tolerant to *P. syringae* infection than WT plants. *AN5_D5* and *AN5_G3* show a significantly lower leaf damage and *P. syringae* biomass but the amount of cell death is not significantly lower although a lower trend can be observed. *AN5_D5* and *AN5_G3* are most probably also more tolerant to *P. syringae* infection than WT plants. The trypan blue staining in mock treated plants was not quantified as there was no cell death at all in the mock treated plants (Figure 4.8).

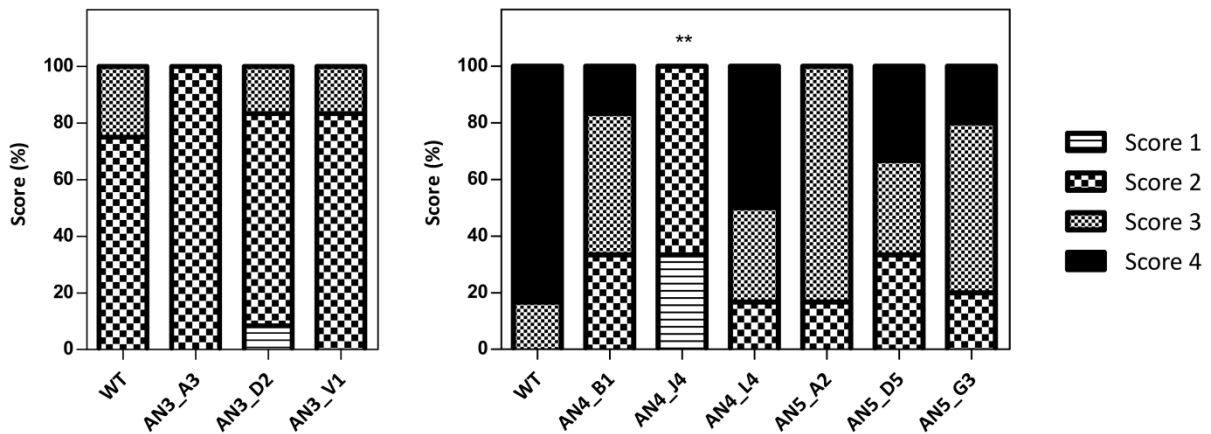


Figure 4.7 Scoring of trypan blue stained leaves from overexpression lines compared to WT plants at 4 dpi. Bars represent the scores in percentage from two independent biological replicates (N = 2) with each time three (35S::AN4 and 35S::AN5) or six leaves (35S::AN3) stained per line. Score 1: 0 %, score 2: 1 – 30 %, score 3: 31 – 60 % and score 4: 61 – 100 % trypan blue staining or cell death. Asterisks indicate statistically significant differences to the trypan blue staining in WT plants (* p≤0.05, ** p≤0.01, ***p≤0.001; Mann-Whitney U test).

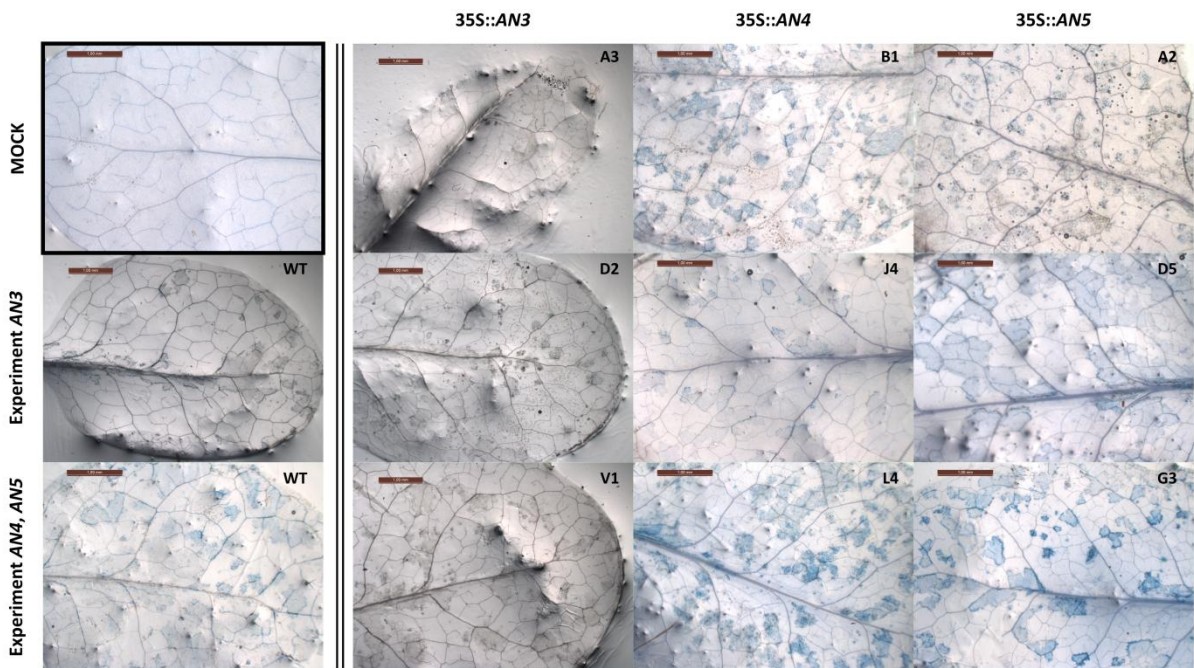


Figure 4.8 Trypan blue stained leaves from WT and overexpression lines at 4 dpi. Representative pictures are shown for each line. Only one picture of a mock treated leaf is shown (black square). Two WT leaves are shown representing different experiments.

4.5 Discussion

A. thaliana ecotype Col-0 was transformed with the overexpression constructs 35S::*AN3*, 35S::*AN4* and 35S::*AN5* and transgenic lines were selected. The resulting homozygous *AN3*, *AN4* and *AN5* overexpression lines carrying a single copy of the T-DNA insertion were tested on transcript level with qRT-PCR using cDNA from 15-day-old *Arabidopsis* seedlings. The overexpression level of the different lines was determined. The validation of the overexpression lines on protein level never showed overexpressed *ArathNictaba* protein although the lanes of the protein gels were overloaded with a total protein amount of 200 µg. Western blot analysis with the anti-Nictaba antibody revealed no indication for overexpression of *ArathNictaba* protein. Since there is no specific antibody for the *ArathNictaba* proteins available, it was worth trying to detect the *ArathNictaba* proteins with the polyclonal anti-Nictaba antibody (raised against Nictaba from tobacco). Western blot analysis with a specific antibody could give decisive answers on the presence of *ArathNictaba* proteins in the overexpression lines.

The expression analysis in chapter 3 revealed an overall downregulation of the expression of *AN4* after salt stress. In contrast, the expression of *AN5* showed a small but significant upregulation after 10 and 24 hours of salt stress. To check whether these *ArathNictaba* genes play a role in the tolerance of the plant against salt stress, the germination of the 35S::*AN4* and 35S::*AN5* transgenic lines was tested on 1/2 MS medium supplemented with different salt concentrations. Whereas two soybean Nictaba-like lectins confer tolerance to high salt stress (150 mM) (Van Holle *et al.*, 2016), the *ArathNictaba* overexpression lines are more sensitive to salt stress than WT plants. It has to be mentioned that the germination data for the WT plants are different from Van Holle *et al.* (2016). Whereas the WT seeds ecotype Col-0 from Van Holle *et al.* (2016) germinated only 64-74 % on 1/2 MS medium supplemented with 150 mM NaCl, a germination of almost 95 % was reached in our experiment. Because of the high percentage of germination for the WT plants, it was expected that none of the overexpression lines would germinate better. Literature searches for comparable experiments revealed only 35 % seed germination on 1/2 MS medium with 150 mM NaCl on day six (Salas-Muñoz *et al.*, 2012). However, Salas-Muñoz *et al.* (2016), a more recent publication of the same research group, reported a germination of 80 % on day four. It can be concluded that not every WT seed batch, even though it concerns the same ecotype tested under similar growth conditions, shows the same germination percentage.

Using *A. thaliana* plants stably expressing an *F-box Nictaba* promoter-β-glucuronidase fusion construct, Stefanowicz *et al.* (2016) showed preferential *F-box Nictaba* promoter activity in trichomes present on young rosette leaves. qRT-PCR analyses verified high expression of *F-box Nictaba* in leaf trichomes. Moreover, overexpression of *F-box Nictaba* resulted in a reduction of leaf damage upon infection with *P. syringae* (Stefanowicz *et al.*, 2016) suggesting a possible function of *F-box Nictaba* and trichomes in plant responses to pathogen attack. Overexpression of the *ArathNictaba* genes revealed a reduced

susceptibility of one of the *AN4* overexpression lines (*AN4_B1*) upon infection with *P. syringae* than WT plants. This result was expected because the expression level of *AN4* showed a late twofold upregulation (5 - 7 dpi) after *P. syringae* infection of WT plants (Chapter 3). Furthermore, transcriptional profiling of mature Arabidopsis trichomes revealed expression of *AN4* in Arabidopsis trichomes (Jakoby *et al.*, 2008). It would be interesting to investigate this tissue specific expression more in detail to determine if there is a link between trichome specific expression and a reduced susceptibility to *P. syringae* infections.

Judging from the *P. syringae* infections of the *AN3* overexpression lines, these lines are not more tolerant to *P. syringae* infections than WT *A. thaliana* plants. This result was unexpected because of the 2 - 2.5 fold upregulation of *AN3* expression after *P. syringae* infections of WT plants 1 - 3 dpi (Chapter 3). Infection experiments with the *AN3* overexpression lines were performed separately from infection experiments with *AN4* and *AN5* overexpression lines. In all experiments, WT plants were included. Between these experiments, important differences can be observed in the percentage of leaf damage and cell death. These percentages are overall much lower in the infection experiments with the *AN3* overexpression lines. Possibly the season in which the infection experiments were performed, is influencing the degree of infection. The infection experiments with *AN4* and *AN5* overexpression lines were performed in spring while the infection experiments with *AN3* overexpression lines were performed in summer. No research showed the direct effect of the season on the degree of *Pseudomonas* infection but Kus *et al.* (2002) revealed the effect of plant age on the degree of *P. syringae* infection. Several times it was observed that Arabidopsis plants grown in growth chambers with strict control of temperature and light regime still sense the season and grow differently in different times of the year (Roden and Ingle, 2009; Bhardwaj *et al.*, 2011; Zhang *et al.*, 2013; Lu *et al.*, 2017). As such, no definitive conclusions concerning the susceptibility of *AN3* overexpression lines compared to WT Arabidopsis plants can be drawn. To draw conclusions, infection experiments have to be repeated with all overexpression lines in the same season.

In contrast with the more or less stable expression of *AN5* after *P. syringae* infection of WT plants of *A. thaliana* (Chapter 3), all *AN5* overexpression lines are most probably more tolerant to *P. syringae* infection than WT plants. Similarly, overexpression of two soybean *Nictaba*-like lectins in *A. thaliana* resulted in less disease symptoms compared to the WT plants (Van Holle *et al.*, 2016). All together, these data showed that the *Nictaba* lectin family might play an important role in the plant defence against *P. syringae* infection.

Chapter 5

Recombinant expression of the Nictaba homologs and interaction study

5.1 Abstract

The expression levels of the ArathNictabas in normal and stressed Arabidopsis plants are too low to allow purification of sufficient amounts of protein starting from plant material. Therefore, *Escherichia coli* was chosen as a heterologous expression system to express the recombinant ArathNictabas and to obtain sufficient amounts of protein. In this chapter, the recombinantly produced AN4 protein was used to determine lectin activity and carbohydrate specificity. Moreover, possible interaction partners for AN4 were searched to improve understanding of the functions and biological importance of the ArathNictabas. In the agglutination assay, no lectin activity could be proven for the purified AN4. MS analysis of the proteins retrieved from the pull-down analysis identified two possible interaction partners, namely TGG1 or myrosinase 1 and BGLU23/PYK10 or β -glucosidase 23. Similar to AN4, these two interaction partners play a role in the plant defence.

5.2 Introduction

The first plant lectins discovered were mainly vacuolar lectins and were present in high concentrations especially in seeds and vegetative storage tissues (Peumans and Van Damme, 1995). Their high abundance made it possible to purify these lectins directly from the plant tissues. The ArathNictabas belong to the Nictaba lectin family, which is one of the six families containing nucleocytoplasmic, inducible lectins (Lannoo and Van Damme, 2010). The expression levels of the ArathNictabas in normal and stressed Arabidopsis plants are too low to allow the purification of sufficient amounts of protein starting from plant material (Chapter 3), similarly to Nictaba from tobacco (Lannoo *et al.*, 2007*a,b*). Therefore, heterologous recombinant expression of these proteins is the only option to obtain sufficient amounts of protein.

Several microorganisms are used as a host to produce recombinant proteins including bacteria, yeast, filamentous fungi and unicellular algae. Also mammalian, plant or insect cell cultures, transgenic plants and animals are possible host systems for recombinant protein production (Demain and Vaishnav, 2009; Rosano and Ceccarelli, 2014). All systems have their strengths and weaknesses, and therefore the choice is dependent on the protein of interest (Demain and Vaishnav, 2009). For example, prokaryotic expression systems may not be suitable if eukaryotic post-translational modifications are necessary (Sahdev *et al.*, 2008). Still, *E. coli* is the least expensive, easiest and fastest expression system to produce recombinant proteins (Demain and Vaishnav, 2009) and possesses several advantages: fast growth kinetics, easy production of high-cell density cultures, inexpensive carbon source, easy transformation and simple to scale up (Sahdev *et al.*, 2008; Rosano and Ceccarelli, 2014).

The development of MS and large-scale proteome analyses created an increasing number of identified proteins (Fukao, 2012). Hardly any protein operates alone while performing its function *in vivo*. Indeed, over 80 % of the proteins have a function in protein complexes (Rao *et al.*, 2014). Protein-protein interactions are crucial for all biological processes ranging from the formation of cellular structures and enzymatic complexes to the regulation of signalling pathways (Lalonde *et al.*, 2008). Studying protein-protein interactions provides fundamental insights into gene function and complex cellular networks (Zhang *et al.*, 2010a; Fukao, 2012). To date, various methodologies to determine protein-protein interactions are available including the yeast two-hybrid system, affinity purification and MS analyses, surface plasmon resonance spectroscopy, nuclear magnetic resonance, fluorescence imaging, *in silico* prediction and protein microarrays. All these techniques have their advantages and limitations (Lalonde *et al.*, 2008; Zhang *et al.*, 2010a; Fukao, 2012; Rao *et al.*, 2014).

The Arabidopsis proteome can be compared with an iceberg. The physiological function of 90 % of the Arabidopsis genes (the hidden part of the iceberg) is unknown, although for two third of these genes, ideas for molecular function can be deduced from sequence similarity. Only a tiny fraction of the genes (the tip of the iceberg) are experimentally and functionally characterized in detail. Protein-protein interactions between proteins of the tip and the hidden part of the iceberg will gain insights into the functions of unknown proteins (Braun *et al.*, 2013).

Glycan array analysis revealed specificity of Nictaba from tobacco for GlcNAc oligomers, high mannose and complex *N*-glycans (Lannoo *et al.*, 2006). Moreover, recombinant Nictaba expressed in the yeast *Pichia pastoris* demonstrated lectin activity in an agglutination assay (Lannoo *et al.*, 2007b). Interaction of Nictaba with several core histones via *O*-GlcNAc was revealed using lectin affinity chromatography and pull-down assays (Schouppe *et al.*, 2011).

In this chapter, recombinant AN4 protein was produced in *E. coli* and used to determine lectin activity and carbohydrate specificity. Moreover, interaction partners were searched for AN4 to improve our understanding of the functions and biological importance of Nictaba homologs from *A. thaliana*. Recombinant protein production was attempted for AN3, AN4 and AN5 in two eukaryotic expression systems namely *Pichia pastoris*, a yeast, and bright yellow-2 (BY-2) cells, a tobacco cell culture. With both expression systems, no recombinant protein could be detected and purified for either of the ArathNictabas. Recombinant protein production in *E. coli* resulted in a small amount of AN4 protein, while no protein for AN3 and AN5 could be detected in the soluble fraction.

5.3 Materials and methods

5.3.1 Plant material and growth conditions

WT *A. thaliana* seeds, ecotype Col-0, were purchased from Lehle Seeds (Round Rock, Texas, USA). After a 3 days stratification period at 4 °C in the dark, Arabidopsis seeds were grown *in vitro* at 21 °C in a controlled growth chamber with a 16/8 h photoperiod. All plant materials and growth conditions were previously described in chapter 3, section 3.3.1.

5.3.2 Construction of *ArathNictaba* His6-tagged constructs

Gibson assembly was used to assemble the N- and C-terminally His6-tagged *ArathNictaba* sequences with the pET-21a(+) vector (Figure 5.1) (Novagen) (Gibson *et al.*, 2009). Gibson assembly can assemble multiple overlapping DNA molecules using a combination of 5' exonuclease, DNA polymerase and DNA ligase (Figure 5.2) (Gibson *et al.*, 2009).

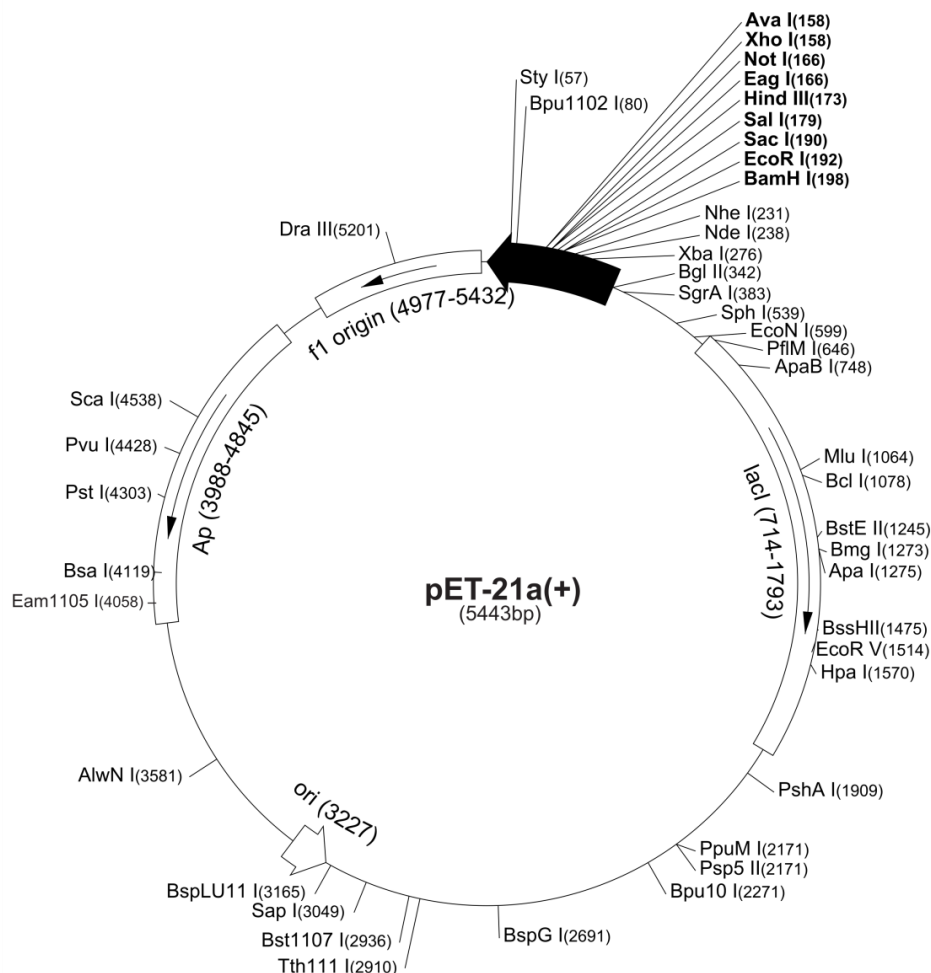


Figure 5.1 Vector map of pET-21a(+) vector (Novagen). The most important parts within the black arrow are the T7 promoter, lac operator, multiple cloning site and T7 terminator. Ap: gene encoding ampicillin resistance; ori: origin of replication in *E. coli*.

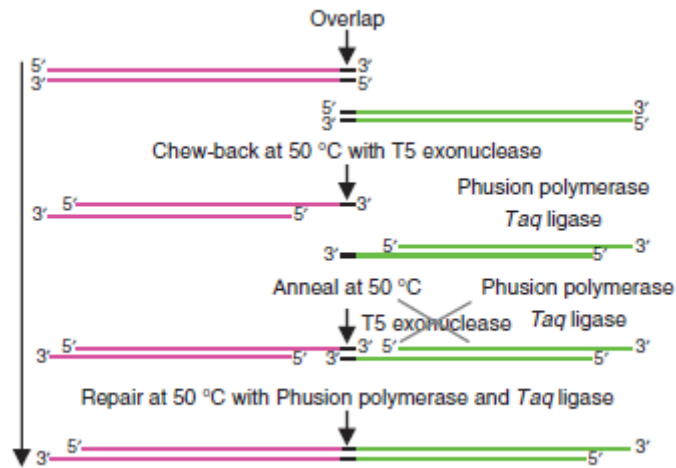


Figure 5.2 Gibson assembly mechanism (Gibson *et al.*, 2009). First, T5 exonuclease removes nucleotides from the 5' ends of two adjacent doublestranded DNA molecules (magenta and green). Second, complementary single-stranded DNA overhangs anneal and T5 exonuclease is inactivated during the 50 °C incubation. Third, Phusion DNA polymerase fills the gaps and Taq DNA ligase seals the nicks (Gibson *et al.*, 2009).

The overlapping parts between the DNA molecules, needed to be assembled, are called the Gibson assembly sites. Both N- and C-terminal His6-tagged ArathNictabas were provided with Gibson assembly sites using PCR. For the N-terminal His6-tagged ArathNictabas, the open reading frame was used as a template for the first PCR with a forward primer to add the N-terminal His6-tag and Gly3-linker, and a reverse primer to add the Gibson assembly site (Supplementary table 8). The second PCR was performed on the product of the first PCR, using a forward primer to add the Gibson assembly site and the same reverse primer as in the first PCR (Supplementary table 8). For the C-terminal His6-tagged ArathNictabas, the open reading frame followed by a Gly3-linker and a His6-tag was used as a template for PCR. Forward and reverse primers were used to add Gibson assembly sites (Supplementary table 8). The vector backbone of the pET21a vector was amplified in a PCR reaction with the following components: 5 x Q5 reaction buffer, 2 mM dNTP mix, 0.1 - 1 ng pET21a, Q5 high fidelity DNA polymerase, water, forward and reverse primer (10 μ M) (Supplementary table 8). Linearization of the vector backbone was performed by *DpnI* restriction (1 hour at 37 °C).

Once the constructs and the vector backbone both contain the Gibson assembly sites, the PCR products were purified with the InnuPREP PCR pure kit (Analytik Jena, Germany). For the Gibson assembly reaction, equimolar amounts of linear vector backbone and ArathNictaba expression construct were mixed together with 15 μ L of Gibson master mix (Table 5.1 and Table 5.2) and incubated for one hour at 50 °C.

Table 5.1 Gibson master mix (25 Gibson assembly reactions).

Component	μL	Manufacturer
5 x ISO buffer	100	/
T5 exonuclease (10000 U/mL)	0.2	NEB
Q5 high fidelity polymerase (10000 U/mL)	6.25	NEB
Taq DNA ligase (40000 U/mL)	50	NEB
MilliQ water	218.55	/

Table 5.2 Components of 5 x ISO buffer.

Component	μL	Manufacturer
1 M Tris-HCl pH 7.5	3000	Sigma Aldrich
2 M MgCl_2	150	VWR
100 mM dNTP mix	240	Thermo Fisher
1 M DTT	300	Sigma-Aldrich
100 mM NAD	300	Sigma Aldrich
PEG-8000	1.5 g	Sigma-Aldrich
MilliQ water	until 6000	/

5.3.3 Transformation of *E. coli* and expression analysis

Half of the Gibson assembly mixture was transformed into heat shock competent *E. coli* strain Rosetta(DE3) cells (Novagen). The strain background is *E. coli* B and Rosetta(DE3) is derived from BL21 lacZY (TunerTM). The strain contains rare codon tRNAs which provide tRNAs for mammalian codons that rarely occur in *E. coli*. After heat shock, transformants were grown on LB agar plates supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin and 25 $\mu\text{g}/\text{mL}$ chloramphenicol. Transformants were subsequently screened by colony PCR with primers that contain the Gibson assembly sites (Supplementary table 8). The pET21a plasmids containing the His6-tagged ArathNictabas were purified using the GeneJET Plasmid Miniprep kit (Life Technologies) and sequenced by LGC Genomics (Berlin, Germany) with a forward sequencing primer on pET21a (Supplementary table 8).

Recombinant *E. coli* Rosetta(DE3) cells were grown overnight in 5 mL LB supplemented with 20 $\mu\text{g}/\text{mL}$ ampicillin and 25 $\mu\text{g}/\text{mL}$ chloramphenicol at 37 °C on a rotary shaker (185 rpm). The next morning, the *E. coli* cells were diluted in 50 mL LB supplemented with 200 $\mu\text{g}/\text{mL}$ carbenicillin and 25 $\mu\text{g}/\text{mL}$ chloramphenicol, and grown at 30 °C on a rotary shaker (185 rpm). After the cells reached an $\text{OD}_{600\text{nm}}$ of 0.5 - 0.6, the expression of recombinant protein was induced with isopropyl β -D-1-thiogalactopyranoside (IPTG). The concentration of IPTG was optimized, IPTG concentrations ranging from 0.1 to 1 mM were checked. After application of IPTG, the *E. coli* cells were grown overnight at 25 °C on a rotary shaker (185 rpm).

The next day, the *E. coli* cell cultures were harvested by centrifugation at 8000 rpm for 15 min. The cell pellets were kept overnight at -20 °C and afterwards solubilised in 10 mL of 1 x phosphate buffer (PB; 0.02 M NaH₂PO₄·2H₂O, 0.23 M Na₂HPO₄) with 500 mM NaCl and 1 mg/mL lysosyme at pH 8. Cell lysis will already occur by freezing the cells and incubation in the presence of lysozyme which causes enzymatic lysis. Additionally, the cells were sonicated three times for 2.5 min. After sonication, the solutions were centrifuged at 4 °C for 45 min (9000 rpm) to separate the soluble from the insoluble fractions. The insoluble fractions were resuspended in 8 M ureum. Both fractions were checked for the presence of recombinant protein by SDS-PAGE and Western blot analysis (section 5.3.6).

5.3.4 Optimization of recombinant protein expression in *E. coli*

To optimize the expression of recombinant protein in the soluble fraction, different culture conditions were tested. The IPTG concentration can be decreased, several IPTG concentrations were tested in the range of 0.1 to 1 mM. 0.2 mM proved to be the best IPTG concentration for optimal protein expression and was used in further experiments. The growth temperature after induction with IPTG can also be diminished, temperatures between 14 and 25 °C were tested. Growth after induction at 14 °C proved to be the best option. When growing bacterial cultures in large volumes (300 mL), the time of induction was also prolonged to three days to obtain more recombinant protein.

5.3.5 Protein purification using column chromatography

E. coli cell extracts containing the recombinant ArathNictaba were obtained as described in section 5.3.4 and used for further purification. 10 mM imidazole (IZ) was added to the soluble protein solution to diminish aspecific protein binding and the pH was adjusted to pH 8. This protein solution was loaded on a Ni-NTA agarose column (MCLAB, South San Francisco, California) equilibrated with 1 x PB, 1 M NaCl at pH 8. Histidine tags have a high affinity for nickel ions, as a result the His₆-tagged recombinant protein will be retained on the column. The column was washed with 1 x PB, 1 M NaCl, 50 mM IZ, pH 8 and 1 x PB, 1 M NaCl, 75 mM IZ, pH 8 to remove aspecific proteins. Elution of the column was performed with 1 x PB, 1 M NaCl, pH 8 with increasing concentrations of IZ ranging from 100 to 500 mM IZ. Several elution fractions of 500 µL were collected for each elution buffer. The OD_{280nm} of each fraction was measured with the Nanodrop 2000 spectrophotometer (Thermo Scientific). The purity of the protein samples was verified by SDS-PAGE and Western blot analysis (section 5.3.6).

5.3.6 SDS-PAGE and Western blot

Protein samples were analyzed by SDS-PAGE on 15 % acrylamide gels as described by Laemmli (1970). After separation, proteins were visualised by gel staining with Coomassie Brilliant Blue R-250 or blotted onto polyvinylidene fluoride transfer membranes (FluoroTrans® PVDF, Pall Laboratory, USA). Membranes were blocked with Tris-buffered saline (TBS; 10 mM Tris, 150 mM NaCl, 0.1 % (v/v) Triton X-100, pH 7.6) containing 5 % (w/v)

non-fat milk powder. Subsequently, membranes were incubated for 1 hour with a mouse monoclonal anti-His6 antibody (Thermo Fisher Scientific) diluted 1/3000 in TBS. After washing three times with TBS, membranes were incubated for 1 hour with the 1/1000 diluted rabbit anti-mouse IgG secondary antibody labelled with horseradish peroxidase (Dako, Glostrup, Denmark). After washing two times with TBS and one time with 0.1 M Tris-HCl buffer (pH 7.6), immunodetection was achieved using a colorimetric assay with 0.1 M Tris-HCl buffer (pH 7.6) containing 700 μ M 3,3'-diaminobenzidine tetrahydrochloride (Sigma-Aldrich, St Louis, USA) and 0.03 % (v/v) hydrogen peroxide. The detection reaction was stopped after 2 - 10 minutes by washing the membrane with distilled water. All washes and incubations were performed on a gently shaking platform at room temperature.

5.3.7 Agglutination assay

To check for lectin activity in purified recombinant protein fractions, agglutination assays were performed as described by Al Atalah *et al.* (2011). Lectins can agglutinate rabbit erythrocytes by binding to the carbohydrate structures on the lipids and proteins in the plasma membrane of the erythrocytes, as such cross-linking the rabbit erythrocytes which results in the formation of a precipitate of the lectin-erythrocyte complexes. Purified protein fractions (10 μ L) were mixed with 10 μ L 1 M ammonium sulfate and 30 μ L of a 20 % solution (1 x PB) of trypsin-treated rabbit red blood cells (Bio-Mérieux, Marcy l'Etoile, France). At different timepoints, agglutination was checked and samples that yielded no visible agglutination activity after 1 hour were noted as lectin negative.

5.3.8 Affinity chromatography with carbohydrates and glycoproteins

Affinity matrices with immobilized fetuin, GlcNAc, D-galactose, D-mannose and ovomucoid were tested for binding with the purified recombinant protein. Approximately 100 μ L of each matrix was used and equilibrated with 1 x PB, 0.5 M NaCl at pH 8 in an Eppendorf tube. The matrix was centrifuged for 1 minute at 4000 g and the equilibration buffer was removed. Afterwards, 200 μ L of the elution fractions containing purified recombinant protein (from the Ni column chromatography) was incubated with each matrix for 2 hours on a turning wheel. After centrifugation and removal of the protein sample, the matrix was washed with 1 x PB, 0.5 M NaCl at pH 8. Finally, after removing the wash solution, all matrices were eluted with 20 mM 1,3-diaminopropane. Run through, wash and elution fractions were evaluated with Western blot analysis (section 5.3.6).

5.3.9 Pull-down analysis

The pull-down analysis was performed directly on the AN4 protein bound to the Ni matrix. It was not possible to work with purified and eluted recombinant AN4 because the IZ could not be removed from the purified protein fractions. Attempts to remove IZ through dialysis resulted in precipitation of the protein and as such loss of the recombinant AN4 protein. Moreover, ZebaTM desalting spin columns (molecular weight cut-off: 7 kDa, Thermo Scientific) to remove salt and IZ resulted in loss of two thirds of the purified protein.

In the pull-down assays, fractions from induced and non-induced *E. coli* Rosetta(DE3) cultures containing the C-terminally His6-tagged AN4 (AN4-HIS) pET expression vector were compared. The induced cultures contain the recombinant protein, the non-induced cultures only contain leaky recombinant protein. Furthermore, two different plant lysates were used, plant extracts from 16-day-old *Arabidopsis* plants subjected or not subjected to 150 mM NaCl for 5 hours. The salt stress was applied as described in chapter 3, section 3.3.4. For the pull-down assays, four small columns containing 100 μ L Ni matrix were used. The soluble fractions of the two different *E. coli* cultures, incubated with the Ni matrix, were each combined with the different plant lysates, resulting in four different pull-down analyses (Figure 5.3). Three replicates for each pull-down experiment were performed and analysed with MS.

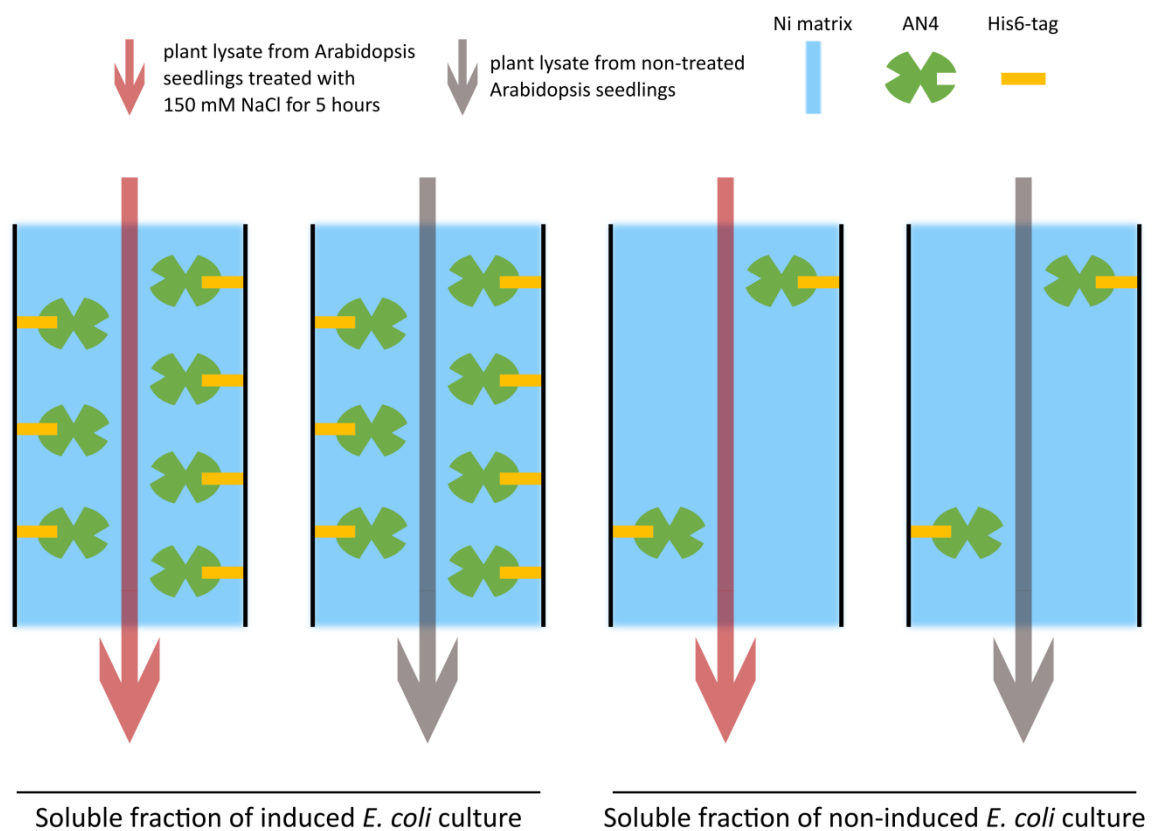


Figure 5.3 Schematic overview of the pull-down experiment. For the pull-down assays, four small columns containing 100 μ L Ni matrix were used. In a first step, the soluble fractions of the two different *E. coli* cultures were incubated with the Ni matrix. Two columns were treated with soluble fraction of an *E. coli* culture that was induced with IPTG to allow AN4-HIS expression, the two remaining columns were treated with a non-induced *E. coli* culture. After washing, the plant lysates from NaCl-treated or control plants were loaded on the columns. Finally, the Ni matrix was washed again and samples were used for trypsin treatment preceding MS analysis.

The preparation of the Ni matrix and the soluble protein solution is described in section 5.3.5. First, the soluble fraction (50 mL) of the *E. coli* culture producing recombinant AN4-HIS was incubated overnight with the equilibrated Ni matrix (100 μ L). The soluble fraction originated from a 300 mL *E. coli* culture induced or non-induced with IPTG to

produce the recombinant AN4 (bait) (section 5.3.3 and 5.3.4). After an overnight incubation, the mixture of Ni-agarose beads and soluble protein solution was transferred in a Poly-Prep purification column (Bio-Rad) and the run through solution was captured. Two wash steps were performed, each until the OD_{280nm} of the captured wash solution was less than 0.05. The first wash buffer consisted of 1 x PB, 1 M NaCl and 50 mM IZ at pH 8, the second wash buffer contained the same components, but with 75 mM IZ. The pH was always fixed to 8 after adding the IZ. After the second wash, the Ni-agarose beads were transferred to an Eppendorf tube with the second wash solution and centrifuged for 2 minutes at 2900 g. After removing the wash solution, 1 mL plant lysate (prey) was added to the beads and this mixture was incubated for 30 - 35 minutes on a turning wheel. The plant lysate was extracted from normal or salt-stressed *Arabidopsis* plant material (0.5 mL) with 1 mL extraction buffer (1 x PB, 1 M NaCl, 25 mM IZ, 10 % glycerol, 0.1 % Tween-20, 1 % β -mercaptoethanol and complete™, Mini, EDTA-free Protease Inhibitor Cocktail at pH 8). After centrifugation and removal of the plant lysate, the beads were washed with the first wash buffer containing 50 mM IZ. Finally, the beads were washed three times with 20 mM Tris-HCl, 2 mM $CaCl_2$ at pH 8 and resuspended in 150 μ L of the same buffer to store at -20 °C prior to MS analysis. All steps described above were performed in a cold room at 4 °C.

Before MS, the samples were treated with trypsin (1 μ g) for 4 hours at 37 °C to cleave all proteins from the Ni beads. After removal of the Ni beads, a second trypsin treatment was performed overnight at 37 °C. Trifluoroacetic acid (1 %) was added to deactivate the trypsin and the samples were subsequently desalted. Next, the samples were dried completely and re-dissolved in 2 % acetonitrile and 0.1 % trifluoroacetic acid. Afterwards, the MS analysis was performed with the Q Exactive™ HF Hybrid Quadrupole-Orbitrap™ Mass Spectrometer at the VIB's Proteomics Expertise Center (Center for Medical Biotechnology, VIB, UGhent). Database searches were achieved with the MaxQuant software and statistical analysis with the Perseus software. Both analyses were outsourced to the VIB's Proteomics Expertise Center.

5.3.10 Molecular modelling

The molecular modelling studies were performed by Prof. Pierre Rougé (University of Toulouse, France). Homology modelling of Nictaba from tobacco and one ArathNictaba (AN4) was performed with the YASARA Structure program (Krieger *et al.*, 2002). Different models of Nictaba and AN4 were built from the X-ray coordinates of the carbohydrate-binding module (CBM) of the GH family 10 protein from *Prevotella bryantii* B14 (PDB code 4MGQ) and *Bacteroides intestinalis* (PDB code 4QPW) (Zhang *et al.*, 2014), and the CBM4-2 of the xylanase from *Rhodothermus marinus* (PDB code 1K42) (Simpson *et al.*, 2002), used as templates. Finally, a hybrid model of the proteins was built using the different previous models. PROCHECK was used to assess the geometric quality of the three-dimensional models (Laskowski *et al.*, 1993). In this respect, all residues of the Nictaba model were correctly assigned in the allowed regions of the Ramachandran plot except for

three residues (Glu2, Pro71 and Arg112), which occur in the non-allowed region of the plot. Similarly, Arg165 of AN4 was found to occur in the non-allowed region of the Ramachandran plot. Using ANOLEA to evaluate the models, only one residue of Nictaba over 165 aa and eleven residues of AN4 over 180 aa exhibited an energy higher than the threshold value (Melo and Feytmans, 1998). The residues were mainly located in the loop regions connecting the β -sheets in the models. The calculated QMEAN6 score of Nictaba and AN4 is 0.36 and 0.38, respectively (Arnold *et al.*, 2006; Benkert *et al.*, 2011). Docking of chitotriose to Nictaba and AN4 was performed with SwissDock (Grosdidier *et al.*, 2011). Molecular cartoons were drawn with the UCSF Chimera package (Pettersen *et al.*, 2004).

5.4 Results

At the start of this PhD the idea was to obtain recombinant protein for AN3, AN4 and AN5. Since AN3, AN4 and AN5 originate from the plant *A. thaliana*, an eukaryotic expression system is the best choice. Two eukaryotic expression systems namely *Pichia pastoris*, a yeast, and bright yellow-2 (BY-2) cells, a tobacco cell culture, were attempted. However, with both expression systems, no recombinant protein for either of the ArathNictabas could be detected or purified. Eventually production of recombinant protein was tested in *E. coli*. However, all efforts to get recombinant protein for the full sequence of AN3 and AN5, and the Nictaba domain of AN3 with a C-terminal His6-tag did not yield any recombinant protein in the soluble fraction of the transformed *E. coli* cultures. Despite many different experiments and culture conditions, the C-terminal His6-tagged protein for AN3, AN5 and the Nictaba domain of AN3 was always detected in the insoluble fraction of the transformed *E. coli* Rosetta(DE3) cultures.

In the case of N-terminally His6-tagged AN4 (HIS-AN4) and AN4-HIS, most recombinant protein was present in the insoluble fraction of the transformed *E. coli* Rosetta(DE3) cells. Fortunately, a small amount of AN4-HIS protein was also present in the soluble fraction. First, it was tried to re-solubilise and refold the AN4-HIS protein out of the insoluble fraction of the transformed *E. coli* Rosetta(DE3) cells. All these attempts resulted in aggregation and precipitation of most of the protein resulting in too low protein yield. Because of these problems, the AN4-HIS protein was purified from the soluble *E. coli* fraction. The results part of this chapter only describes the purification and analyses of recombinant AN4 from the soluble fraction of *E. coli* Rosetta(DE3) cells.

5.4.1 Recombinant expression and purification of AN4 in *E. coli*

To analyze the carbohydrate binding specificity and to identify possible interaction partners, recombinant AN4 was produced in *E. coli* Rosetta(DE3) cells. HIS-AN4 and AN4-HIS constructs were created and transformed in *E. coli* Rosetta(DE3) cells. The expression of the recombinant proteins was induced with 0.2 mM IPTG for 24 hours at 14 °C. The protein samples (10 μ L) analysed with SDS-PAGE and Western blot analysis (Figure 5.4) are derived

from 50 mL *E. coli* Rosetta(DE3) cultures. The calculated size of HIS-AN4 and AN4-HIS is 21.3 kDa. Only a very small amount of HIS-AN4 protein was detected in the insoluble fraction from the induced *E. coli* culture using the anti-His antibody in the Western blot analysis (Figure 5.4C lane 3). On the coomassie stained gel, this protein is not clearly visible (Figure 5.4A lane 3). In contrast, AN4-HIS is clearly visible in the insoluble fraction of the induced *E. coli* culture (Figure 5.4A lane 6). Western blot analysis confirmed that AN4-HIS is mostly present in the insoluble fraction of the induced *E. coli* culture (Figure 5.4C lane 6). However, AN4-HIS is also present, though in lesser amount, in the soluble fraction of the induced *E. coli* culture (Figure 5.4B lane 2, Figure 5.4C lane 7). Furthermore, the soluble and insoluble fractions of the non-induced *E. coli* culture reveal leaky expression of AN4-HIS in the absence of IPTG treatment (Figure 5.4C lane 8, 9). No leaky expression was detected in the case of HIS-AN4 in the absence of IPTG treatment (Figure 5.4C lane 1,2).

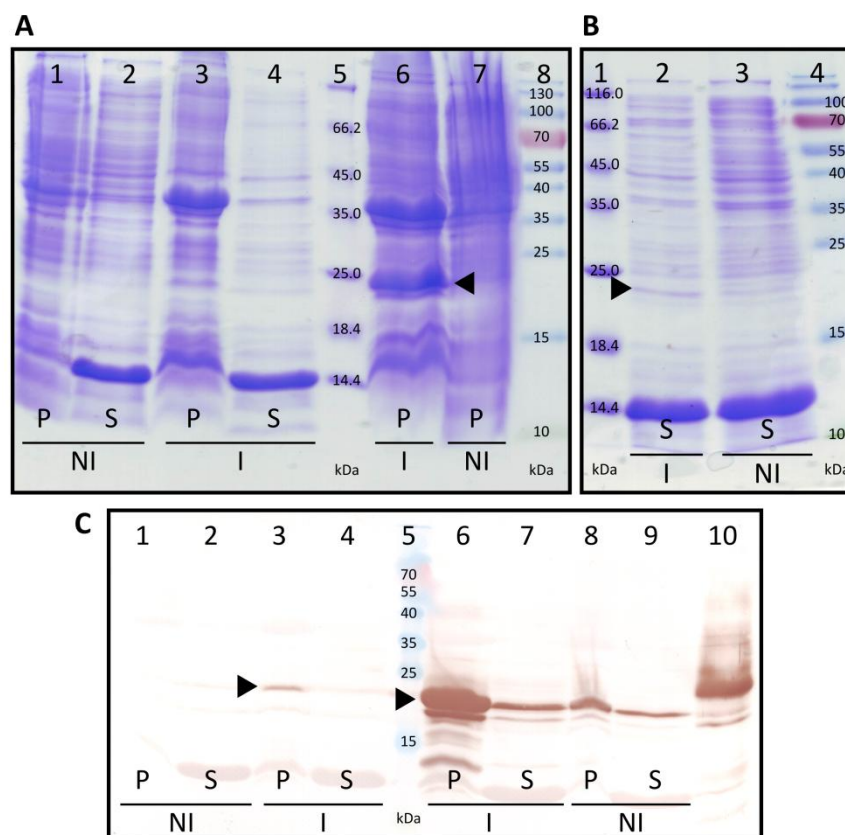


Figure 5.4 SDS-PAGE (A,B) and Western blot (C) analysis of recombinant AN4 produced in *E. coli*. Crude extracts (10 μ L) of the soluble and insoluble fractions of a 50 mL Rosetta(DE3) *E. coli* culture. P: pellet (insoluble fraction), S: supernatans (soluble fraction), I: induced with IPTG, NI: non-induced. The arrows indicate the position of recombinant AN4. **(A)** Lanes 1 - 4: HIS-AN4; lane 5: protein marker (Thermo Fisher Scientific); lane 6 - 7: AN4-HIS; lane 8: prestained protein marker (Thermo Fisher Scientific). **(B)** Lane 1: protein marker; lanes 2 - 3: AN4-HIS; lane 4: prestained protein marker. **(C)** Lanes 1 - 4: HIS-AN4; lane 5: prestained protein marker; lanes 6 - 9: AN4-HIS; lane 10: positive control (*Oryza sativa* EULS2).

Next, the expression of the AN4-HIS protein was optimized to get as much protein as possible in the soluble fraction of the *E. coli* Rosetta(DE3) cells. Induction of the 300 mL bacterial cultures with 0.2 mM IPTG for three days at 14 °C proved to be the best condition. Purification starting from the soluble fraction was performed with Ni-NTA agarose beads. To enhance extraction and binding of AN4-HIS to the matrix, experiments were performed in the presence of different concentrations of salt, either 1 x PB buffer containing 0.5 M NaCl or the same buffer containing 1 M NaCl, both buffers fixed at pH 8. Protein fractions obtained after nickel affinity chromatography were analysed with SDS-PAGE and Western blot analysis (Figure 5.5). AN4-HIS is present in all tested elution fractions, although the amount of protein and the purity differs depending on the salt concentration used during purification. A comparison between the elution fractions obtained with 250 mM IZ for the purifications performed in the presence of 0.5 and 1 M NaCl revealed that the ratio between AN4-HIS and contaminants is much better when the purification is done with 1 M NaCl (Figure 5.5A and B coomassie stained gels). For the purification in the presence of 1 M NaCl, an extra wash step with 75 mM IZ and two extra elution steps with varying IZ concentrations were included (Figure 5.5B).

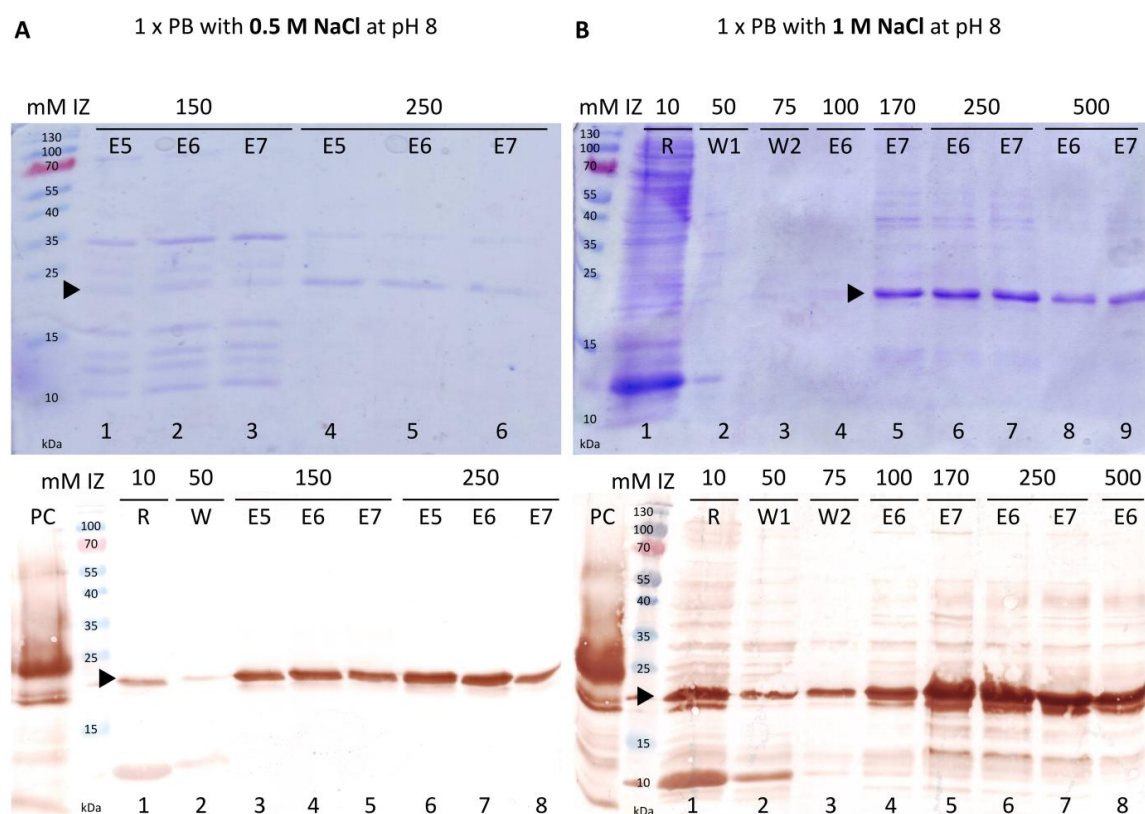


Figure 5.5 SDS-PAGE and Western blot analyses of purified AN4 protein. Purifications were started from a 300 mL *E. coli* Rosetta(DE3) culture expressing AN4-HIS after treatment with 0.2 mM IPTG for three days at 14 °C. The arrows indicate the position of recombinant AN4. **(A)** 1 x PB, 0.5 M NaCl pH 8 was used as extraction, wash and elution buffer. **(B)** 1 x PB, 1 M NaCl pH 8 was used as extraction, wash and elution buffer. IZ concentrations in each buffer are indicated at the top of the gel. Samples (15 μ L) were analyzed on gel. R: run through, W: wash solution, E: elution fractions, PC: positive control (*Oryza sativa* EULS2).

Elution fraction number five (E5; $OD_{280nm} = 0.068$; concentration = 0.030 mg/mL), eluted with 250 mM IZ, from the purification in the presence of 0.5 M NaCl (Figure 5.5A lane 4) was used for agglutination assays. The protein was incubated with trypsin-treated rabbit erythrocytes and incubated at room temperature. Agglutination was checked at different time points, but no visible agglutination activity was observed, not even after 1 hour. In these assays, purified Nictaba from tobacco was used as a positive control and agglutinated the rabbit erythrocytes. The 250 mM IZ elution fractions from the purification in the presence of 0.5 M NaCl (Figure 5.5A lanes 4, 5 and 6) were used to test the binding of recombinant AN4 to different affinity matrices with immobilized carbohydrates and glycoproteins. Fetuin, GlcNAc, D-galactose, D-mannose and ovomucoid matrices were tested and the run through, wash and elution fractions were analysed with Western blot analysis (Figure 5.6).

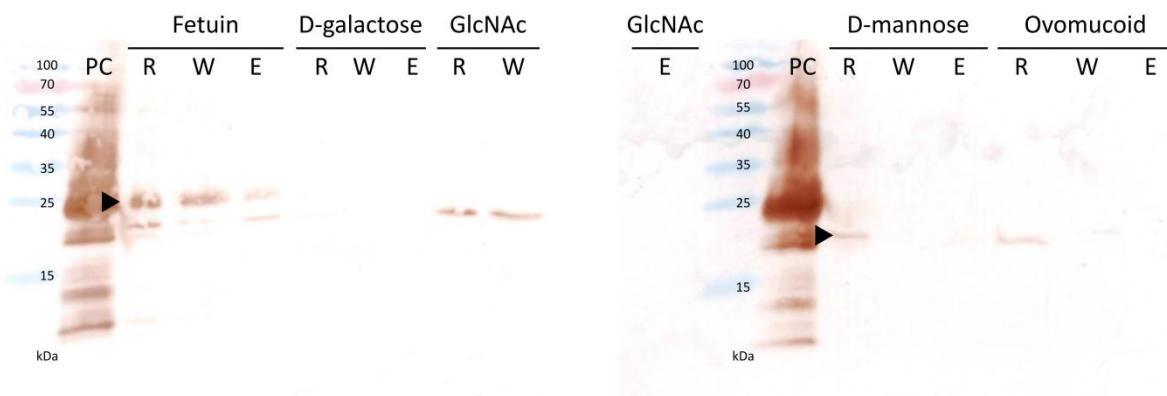


Figure 5.6 Western blot analyses of purified recombinant AN4 protein after affinity chromatography with carbohydrates and glycoproteins. The 250 mM IZ elution fractions from the purification in the presence of 0.5 M NaCl (Figure 5.5A lanes 4, 5 and 6) were used to test the binding of recombinant AN4 to different affinity matrices with immobilized carbohydrates and glycoproteins. Samples (15 μ L) were analyzed on gel. R: run through, W: wash solution, E: elution, PC positive control (*Oryza sativa* EULS2).

The AN4-HIS protein was detected in the run through and/or wash fraction for all matrices, except for the D-galactose matrix. Strangely, for the D-galactose matrix no protein could be detected in any fraction. Clearly, the data show that AN4-HIS could not bind to the fetuin, GlcNAc, D-mannose and ovomucoid matrices. The very faint band in the elution fraction of the fetuin matrix is most probably due to overflow between the lanes.

5.4.2 Pull-down analysis to search for interacting partners of AN4

Purification of recombinant AN4-HIS yielded protein fractions containing AN4 in the presence of 100, 170, 250 and 500 mM IZ. During dialysis to remove IZ, recombinant AN4-HIS showed tendency to precipitate, resulting in loss of the protein. Therefore, the pull-down analysis was performed directly on the AN4 protein bound to the Ni matrix.

A first experiment was performed to test the pull-down assay with induced and non-induced *E. coli* cultures, and plant lysate from non-treated Arabidopsis seedlings. The pull-down assay with the soluble fraction of an induced *E. coli* culture showed a relative abundance of 13.4 % for AN4 whereas the pull-down assay with the soluble fraction of a non-induced *E. coli* culture only showed a relative abundance of 1.2 % (leaky expression).

In the second experiment, the pull-down analysis was also performed with induced and non-induced *E. coli* cultures, but two different Arabidopsis plant lysates were used. One assay was performed with the extract from Arabidopsis seedlings treated for 5 hours with 150 mM NaCl, the other assay used an extract originating from non-treated Arabidopsis seedlings. Furthermore, three replicates of each pull-down were implemented.

In both experiments, the recombinant AN4-HIS from induced and non-induced *E. coli* cultures was first bound to the Ni matrix. Next, a total Arabidopsis plant lysate (extract from non-treated or salt-treated Arabidopsis seedlings) was mixed with the AN4-HIS bound to the Ni beads and as such interaction partners of AN4 could bind (Figure 5.3). After washing, all bound proteins were analysed with MS.

Following pull-down and MS, all *E. coli* and *A. thaliana* proteins were identified according to different databases. The results from the pull-down assays performed with the induced *E. coli* culture were compared to these of the non-induced *E. coli* culture. This comparison excluded the *E. coli* proteins that bind aspecifically to the Ni matrix and the Arabidopsis proteins that bind aspecific to these *E. coli* proteins or to the Ni matrix. Results from the pull-down assays performed with the two different plant lysates, were also compared with each other. The transcript levels of AN4 in Arabidopsis are downregulated after salt stress (Chapter 3, section 3.4.6.2) which can also be the case for the possible interaction partners. Possible interaction partners for which the expression is downregulated by salt stress are expected to be absent in the analysis using the plant lysate from the stressed plants compared with the analysis using the plant lysate from the non-treated Arabidopsis seedlings.

Figure 5.7 shows the volcano plots of the MS analysis of the experiment with three replicates of each pull-down. A volcano plot is a type of scatter plot used to quickly identify significant differences in large datasets and shows significance (p value) versus fold change. Each volcano plot shows a comparison between two datasets, each resulting from three replicates.

The first volcano plot shows the comparison between three pull-down assays performed with the induced *E. coli* culture and three pull-down experiments with the non-induced *E. coli* culture, each using a plant lysate of Arabidopsis seedlings grown under normal conditions (Figure 5.7A). Since all significant hits are *E. coli* proteins and AN4 is not one of the significant hits, the results of this comparison are not useful. AN4 is expected to be a significant hit because it should be present in higher levels in the pull-down assays with the soluble fraction of an induced *E. coli* culture. Analysis of the results for the six individual pull-down experiments revealed that one of the pull-down assays with the soluble fraction of a non-induced *E. coli* culture showed aberrant results from the two other replicates. After MS analysis each individual pull-down experiment resulted in more than 1300 identified proteins, except for this aberrant replicate where only 566 proteins were identified. Since the MS analysis was performed in parallel for all samples, something probably went wrong during the pull-down of this sample or preparation of the sample before MS.

Figure 5.7C shows the results of the same comparison shown in Figure 5.7A but the aberrant replicate was excluded from the results. As expected, AN4 is significantly more present in the pull-down experiments with the induced *E. coli* culture. Next to AN4, three *A. thaliana* proteins originating from the plant lysate are significantly more present in the pull-down experiments with the induced *E. coli* culture. These three proteins, TGG1, ESM1 and PTAC16 are possible interaction partners for AN4. This result has to be interpreted with caution since it relies on two rather than three replicates of the pull-down experiments with the non-induced *E. coli* culture.

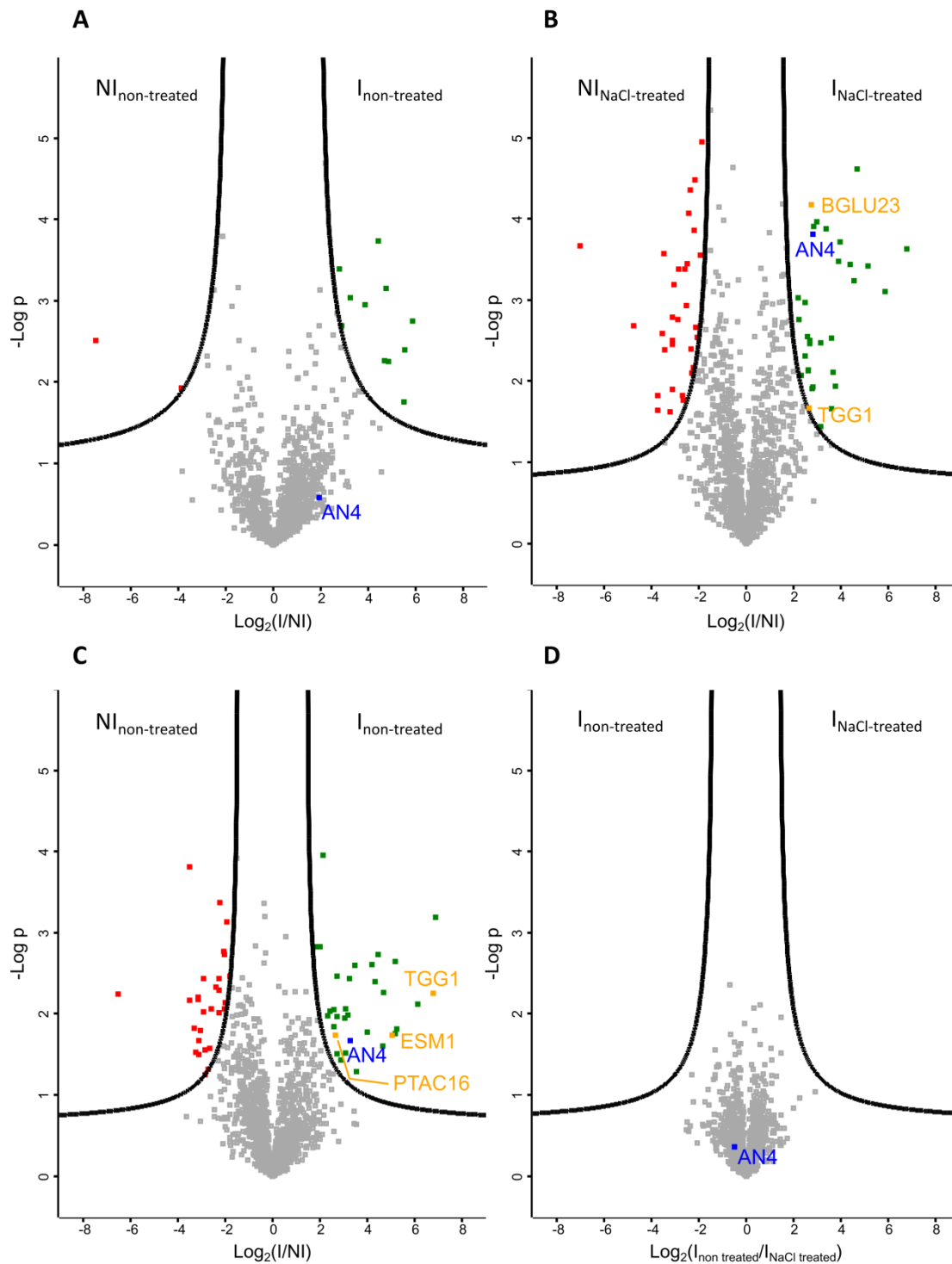


Figure 5.7 Volcano plots MS analysis. (A,C) Plant lysates from non-treated Arabidopsis seedlings. **(B)** Plant lysate from Arabidopsis seedlings treated with 150 mM NaCl for 5 hours. **(D)** Plant lysates from non-treated and treated Arabidopsis seedlings with 150 mM NaCl. AN4 is indicated in blue, possible interaction partners in orange, *E. coli* proteins which are significantly more present in the pull-down assays performed with the induced *E. coli* culture (I) in green and *E. coli* proteins which are significantly less present in these pull-down assays in red. In **plots A, B and D** two datasets were compared (T-test, $p \leq 0.01$), each originating from three replicates. In **plot C**, one replicate of the soluble fraction of a non-induced *E. coli* culture (NI) was excluded and $p \leq 0.05$ was used in the T-test performed on the datasets.

Figure 5.7D compares three pull-down assays with the induced *E. coli* culture using a plant lysate from non-treated Arabidopsis seedlings and three pull-down assays with the induced *E. coli* culture using a plant lysate from salt-treated Arabidopsis seedlings. As expected, AN4 is situated almost in the middle of the volcano plot because AN4 levels should be the same in all soluble fractions of the induced *E. coli* cultures. AN4 levels originating from *A. thaliana* in the plant lysate are probably negligible relative to the recombinantly produced AN4. Since no significant hits are found in this comparison, it can be concluded that the pull-down results with the plant lysate of the non-treated Arabidopsis seedlings are not significantly different from the pull-down results with the plant lysate of the salt stress treated Arabidopsis seedlings.

The volcano plot in Figure 5.7B shows the comparison between three pull-down assays performed with the induced *E. coli* culture and three pull-down assays with the non-induced *E. coli* culture, each using a plant lysate of salt-treated Arabidopsis seedlings. This volcano plot should actually be similar as the volcano plot in Figure 5.7A, but this is not the case because of the aberrant replicate in Figure 5.7A. In Figure 5.7B, AN4 and two *A. thaliana* proteins are significantly more present in the pull-down experiments with the induced *E. coli* culture, concluding these two proteins TGG1 and BGLU23 are possible interaction partners for AN4. TGG1 was also found in Figure 5.7C making it more conclusive to be an interaction partner for AN4. However, BGLU23 has a much lower p value than TGG1 in Figure 5.7B. All significant hits from Figure 5.7A, Figure 5.7B and Figure 5.7C can be found in supplementary tables 9, 10 and 11, respectively.

5.4.3 Molecular modelling of AN4

Three-dimensional models of lectins allow researchers to get new insights in the overall structure of the lectin and in the amino acids that are important in the carbohydrate binding site. The three-dimensional models built for Nictaba and AN4 exhibit both the canonical β -sandwich core structure of the CBM of GH family 10 enzymes (Figure 5.8). The β -sandwich core structures for Nictaba and AN4 superpose nicely (data not shown). However, the size and shape of the loops connecting the strands of the β -sheets are different. Both Trp-residues important for carbohydrate binding of Nictaba are conserved in AN4 and indicated in Figure 5.8.

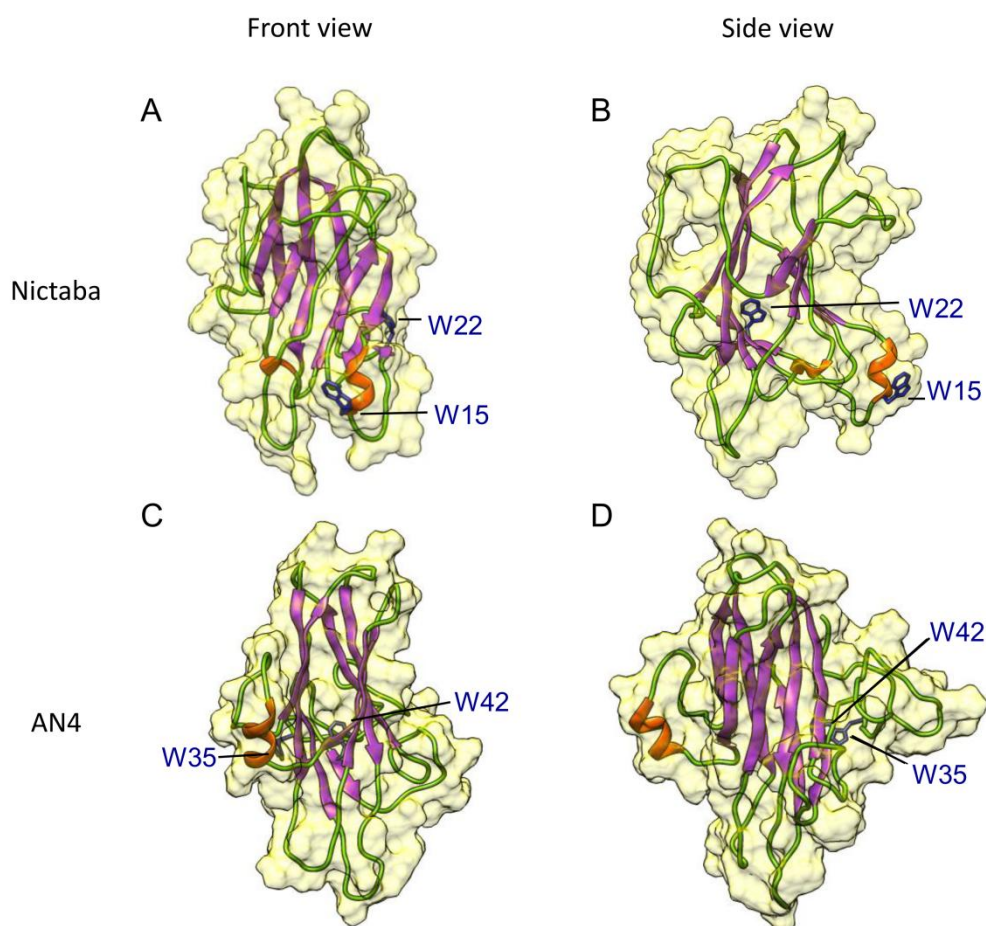


Figure 5.8 Molecular models of Nictaba and AN4. Front (A, C) and side (B, D) view of the ribbon diagrams of the three-dimensional models built for Nictaba and AN4, respectively. The molecular surfaces are coloured pale yellow and the conserved Trp-residues important for the carbohydrate binding of Nictaba are indicated in blue. α -helices, β -sheets and loops/turns are coloured orange, purple and green, respectively. Molecular modelling was performed by Prof. Pierre Rougé.

Docking experiments performed with chitotriose as a ligand, resulted in different carbohydrate binding schemes for Nictaba and AN4. Surface plasmon resonance analyses indicated that chitotriose has the strongest interaction with Nictaba from tobacco (Chen *et al.*, 2002). Three different putative chitotriose-binding sites were identified on the molecular surface of Nictaba and AN4 (Figure 5.9A,C). However, the localization of these putative chitotriose-binding sites is different for Nictaba and AN4. In Nictaba, these binding sites are localized on a single face of the β -sandwich core structure. In AN4, both faces of the β -sandwich structure contribute to the binding of chitotriose. According to the degree of occupancy of the sites by the chitotriose molecules, one of the three chitotriose-binding sites appears as being more crowded than the two others, in both Nictaba and AN4. The more crowded binding site occurs at the convex face of the β -sandwich core structure in Nictaba (Figure 5.9A), whereas it occupies the concave face of the β -sandwich core structure in AN4 (Figure 5.9C). Surprisingly the position of the Trp-residues is not really linked to the localization of the chitotriose-binding sites except for Trp35 in AN4 (Figure 5.9). Although

speculative, these docking results suggest some discrepancies in the carbohydrate binding properties between Nictaba and AN4.

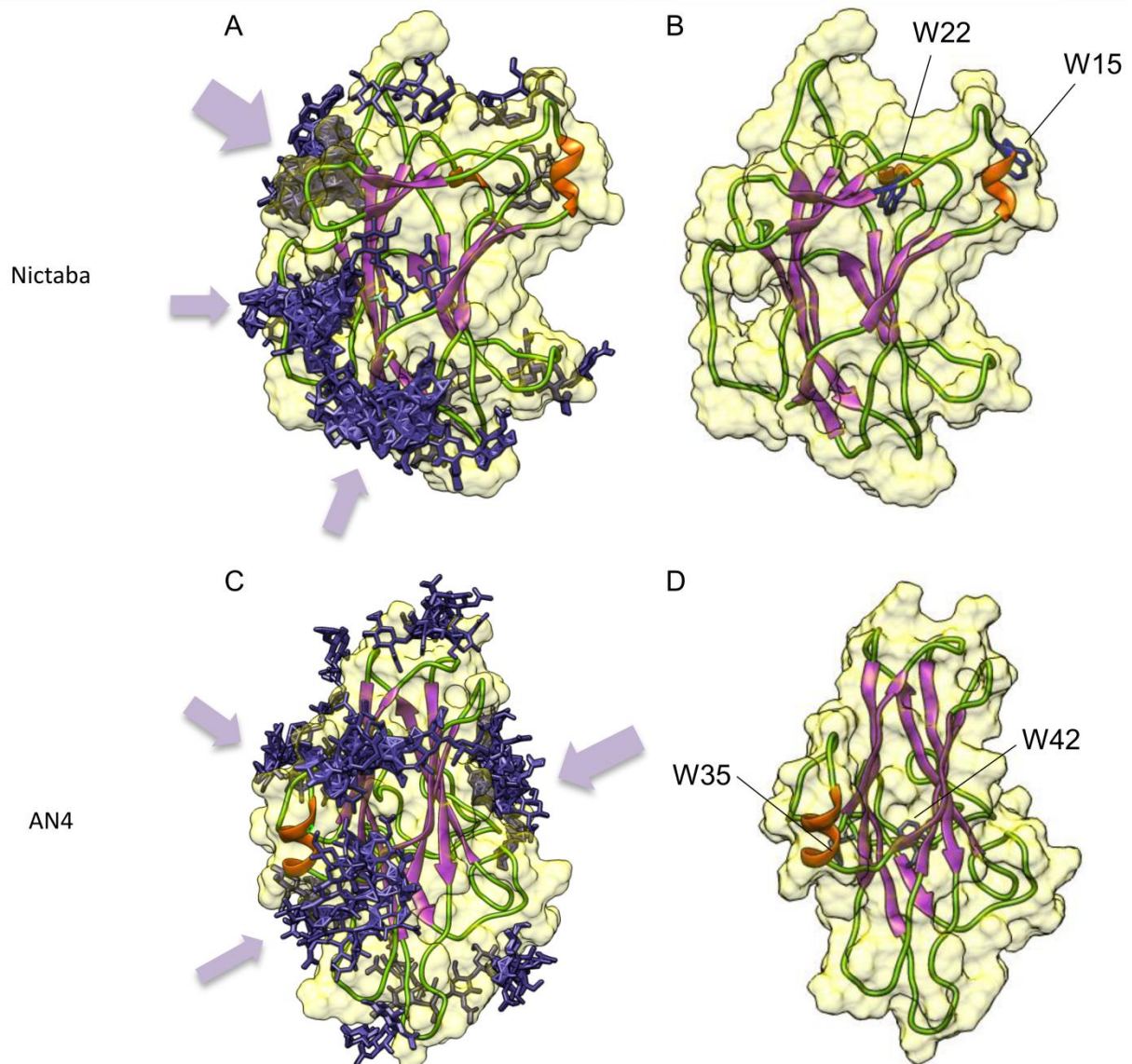


Figure 5.9 Docking of chitotriose to Nictaba and AN4. Ribbon diagrams of the three-dimensional models built for Nictaba (A) and AN4 (C) docked with chitotriose. The arrows indicate the localization of the chitotriose-binding sites. The size of the arrows depends on the degree of occupancy of the putative binding sites by the chitotriose molecules. The conserved Trp-residues important for the carbohydrate binding of Nictaba are indicated in blue for Nictaba (B) and AN4 (D). Docking analyses were performed by Prof. Pierre Rougé.

5.5 Discussion

Since our goal was to express a protein that originates from the plant *A. thaliana*, the choice for an eukaryotic expression system was most obvious. Two eukaryotic expression systems namely *Pichia pastoris*, a yeast, and BY-2 cells, a tobacco cell culture, were tried. No recombinant protein for either of the ArathNictabas could be detected or purified. Eventually, the production of recombinant AN4, both with an N-terminal and C-terminal His6-tag, was tested in *E. coli*. Since AN4 contains no cysteins which can form disulphide bridges and no possible *N*-glycosylation sites, the switch to this prokaryotic expression system should in theory not give too many problems.

Our analyses focused on the expression constructs for AN4 containing a His6-tag at the N- or C-terminal end. Only a very small amount of HIS-AN4 protein in the insoluble fraction of the *E. coli* culture and no HIS-AN4 protein in the soluble fraction was detected (Figure 5.4C). AN4 contains no signal peptide at the N-terminal side, as such interference of the His6-tag with a signal peptide could not be the cause. A buried His6-tag in the fold of AN4 is another possibility (Rosano and Ceccarelli, 2014), but very unlikely because the analysis was performed in denaturing conditions (SDS-PAGE). Another possibility is that HIS-AN4 is not or low expressed, this can be due to a harmful effect of the heterologous protein on the *E. coli* cells (Dumon-Seignovert *et al.*, 2004). The problem of toxicity seems also unlikely since no indication of toxicity related to the expression of AN4-HIS on *E. coli* cell growth was observed. Gustafsson *et al.* (2004) reported codon bias which occurs when the frequency of synonymous codons in the protein that you want to express differs significantly from that of the host. Hereby, certain low abundant tRNAs can be depleted and thus affect heterologous protein expression levels. For example, the AGG codon is a rare codon in *E. coli* (used at frequency of 0.2 %) while in plant mRNA it can reach frequencies of 1.5 % or more (Kane, 1995). The used strain in this thesis is the Rosetta(DE3) strain which contains a plasmid that supplies a lot of tRNAs for rare codons in *E. coli*. Apparently, for HIS-AN4, the usage of this Rosetta(DE3) strain is not resulting in high expression levels. No further optimizations were tried to obtain a higher expression of HIS-AN4 since the *E. coli* cells transformed with the expression construct for AN4-HIS showed a high expression of the protein in the insoluble fraction of the transformed *E. coli* cells. Furthermore, low levels of AN4-HIS protein could also be detected in the soluble fraction of the transformed *E. coli* cells (Figure 5.4C).

In first instance, several trials were performed to re-solubilise and refold the AN4-HIS protein from the insoluble fraction of the *E. coli* cells. Inclusion bodies are formed from misfolded proteins that start to aggregate. For sure when there is a high expression level, hydrophobic stretches in the polypeptides are highly concentrated and thus available for interactions with other hydrophobic stretches (Carrió and Villaverde, 2002). All the attempts to re-solubilise and refold AN4-HIS from the inclusion bodies resulted in aggregation and precipitation of most of the protein. As a result the protein yield was too low for further analyses. Furthermore, re-solubilisation and refolding of proteins from the insoluble fraction is very

labour intensive and error prone without guarantee that the protein will be properly folded and functionally active.

During purification starting from the soluble fraction of the *E. coli* Rosetta(DE3) cells, problems with contaminants were experienced. The use of higher salt concentrations resulted in protein fractions with a higher purity of AN4-HIS and less contaminants. Removal of the IZ and NaCl after purification by dialysis, resulted in precipitation of the protein. Because of this problem, pull-down analysis was performed directly on the Ni matrix.

Agglutination assays with an elution fraction containing AN4-HIS showed no agglutination activity towards rabbit erythrocytes. Lectins generally occur as multimers and as such contain more than one carbohydrate binding site which makes it possible to agglutinate red blood cells (Adamová *et al.*, 2014). One reason to explain the negative agglutination activity of AN4 is because this protein is occurring as a monomer. Another possibility can be that the production of recombinant AN4-HIS protein in *E. coli* didn't yield a functional protein or the protein concentration tested in the agglutination assay was too low. Therefore, binding of AN4-HIS to several carbohydrates and glycoproteins (fetuin, GlcNAc, D-galactose, D-mannose and ovomucoid) was evaluated with elution fractions containing AN4-HIS. For all matrices except D-galactose, AN4-HIS was detected in the run-through fraction indicating that AN4-HIS did not bind to these carbohydrates and glycoproteins. Hence, it is possible that the recombinant AN4-HIS produced in *E. coli* was not a functional protein. However, it cannot be excluded that the IZ and NaCl in the elution fractions interfered with the agglutination activity and the binding to the sugar and glycoprotein matrices. Furthermore it is possible that AN4 preferentially recognizes carbohydrate structures that were not included in our analysis. More experiments are necessary to conclude whether AN4 is a functional lectin and if so, to unravel its carbohydrate specificity.

The purified recombinant AN4 was used to look for binding partners of the protein. Pull-down experiments followed by MS analysis revealed several possible interaction partners for AN4. TGG1 (AT5G26000) or myrosinase 1 is a thioglucoside glucohydrolase localized in the vacuole. BGLU23 (AT3G09260) or PYK10 is a β -O-glucosidase localized in the ER bodies. Next to these two interaction partners that were identified in the pull-down assays with the plant lysate originating from the salt stressed *Arabidopsis* seedlings, three possible interaction partners were found from the analysis with the plant lysate originating from non-treated *Arabidopsis* seedlings. It should be mentioned that in this analysis, one replicate of the soluble fraction of the non-induced *E. coli* cultures was missing. One of these interaction partners is TGG1 which was also a significant hit in the first comparison. GDSL esterase/lipase ESM1 (AT3G14210) and protein plastid transcriptionally active 16 (PTAC16; AT3G46780) are the other two possible interaction partners (Supplementary table 11). ESM1 is a secreted protein with a role in glucosinolate hydrolysis as a myrosinase-associated protein. PTAC16 is a chloroplast protein which regulates the membrane-anchoring functions of the nucleoid (Zhang *et al.*, 2006; Burow *et al.*, 2008; Ingelsson and Vener, 2012). Nothing

decisive can be concluded about the interaction of AN4 with ESM1 and PTAC16 because of the replicate that was missing in this analysis.

TGG1 or BGLU38

Myrosinase 1 belongs to the GH1 family and is a β -thioglucoside glucosidase (Andersson *et al.*, 2009). GHs hydrolyse the glycosidic bond between carbohydrates or a carbohydrate and non-carbohydrate moiety (Ahn *et al.*, 2010). The myrosinases from *A. thaliana*, namely TGG1 - TGG6 are all grouped in the myrosinase gene family (Barth and Jander, 2006). All myrosinases are β -glucosidases and as such TGG1 is also called BGLU38. TGG1 and TGG2 are the most studied of the family and are expressed in leaves, stems and floral organs (Xue *et al.*, 1995). TGG3 is a frame-shifted pseudogene specifically expressed in stamens and petals. TGG4 and TGG5 are root-specific myrosinases and TGG6 is only expressed in flowers (Barth and Jander, 2006; Andersson *et al.*, 2009).

Myrosinases catalyse the hydrolysis of glucosinolates and thereby initiate the formation of isothiocyanates, nitriles, thiocyanates, epithionitriles and other reactive products (Thangstad *et al.*, 2004; Andersson *et al.*, 2009). Glucosinolates or thioglucosides consist of a glucose residue linked to an amino acid derived R-group of aliphatic, aromatic or indole types by a thioglucoside bond (Thangstad *et al.*, 2004). In *A. thaliana*, different glucosinolates with side chains derived from methionine, tryptophan, Phe and isoleucine were found (Barth and Jander, 2006). Myrosinases and glucosinolates are localized in separate plant cells and as such only come in contact with each other upon tissue disruption. Myrosinases are present in the myrosin phloem idioblasts in phloem parenchyma while glucosinolates are present in the S-cells adjacent to the phloem (Thangstad *et al.*, 2004; Barth and Jander, 2006; Andersson *et al.*, 2009). Only upon pathogen invasion or insect herbivory, the myrosinases can degrade glucosinolates and as such produce toxic products to protect the plant against these invaders. Thus, myrosinases play a role in plant defence against microbes and herbivores (Barth and Jander, 2006). TGG1 is also localized in the guard cells, next to the myrosin phloem idioblasts (Thangstad *et al.*, 2004). Islam *et al.* (2009) showed that double *Arabidopsis* mutants for TGG1 and TGG2 are not able to induce stomatal closure upon ABA, MeJA or ROS production. Single mutants of TGG1 and TGG2 are able to induce stomatal closure, as such it can be concluded that TGG1 and TGG2 redundantly function in stomatal closure (Islam *et al.*, 2009). The redundancy of TGG1 and TGG2 was already shown by Barth and Jander (2006) by investigating glucosinolate breakdown in single and double mutants.

BGLU23 or PYK10

BGLU23 or PYK10 is a β -O-glucosidase that, like myrosinase 1, belongs to the GH1 family (Nagano *et al.*, 2008). In total 47 β -glucosidases have been identified in *A. thaliana*, called BGLU1 - BGLU47, BGLU23 or PYK10 and TGG1 or BGLU38, the two identified interaction partners for AN4, are two of these β -glucosidases. BGLU23 is a root and hypocotyl specific β -glucosidase which has an endoplasmic reticulum (ER) retention signal at its C-terminus

(Nitz *et al.*, 2001; Xu *et al.*, 2004). The BGLU23 gene contains 12 exons and 11 introns, four of these 12 exons are identical in size to the myrosinases TGG1 and TGG2. With the exception of two exons, the others differ only slightly with TGG1 and TGG2, demonstrating that the gene structures of BGLU23, TGG1 and TGG2 are closely related and these genes may be derived from a common ancestor (Nitz *et al.*, 2001). Ahn *et al.* (2010) showed that BGLU23 possesses β -O-glucosidase activity, but no β -S-glucosidase or myrosinase activity.

BGLU23 is known to hydrolyze the natural substrate scopolin and other coumarin glucosides similar in structure to scopolin. Scopolin is one of the most abundant secondary metabolites in the Arabidopsis roots (Ahn *et al.*, 2010). The resulting scopoletin is a fungitoxic compound and can be polymerized by peroxidase in the presence of H₂O₂ (Reigh *et al.*, 1973). This can protect plant cells from the oxidative damage caused by pathogens. BGLU23 is, as predicted by its ER retention signal, localized to ER bodies in the roots, which are not present in rosette leaves under normal growth conditions, but can be induced there by MeJA and wounding (Nagano *et al.*, 2005; Ahn *et al.*, 2010). Moreover, BGLU23 is thought to restrict root colonization by *Piriformospora indica*, an endophytic fungus, resulting in repression of defence responses and upregulation of responses leading to a mutualistic interaction (Sherameti *et al.*, 2008). These data show that BGLU23 is induced by both abiotic and biotic stresses. However, also suppression of BGLU23 expression by NaCl and mannitol was shown. Upon disruption of cells, BGLU23 forms large complexes with its binding proteins. One of these binding proteins is PBP1, which is a jacalin homolog from *A. thaliana* (Chapter 2) (Ahn *et al.*, 2010). BGLU23 is glycosylated with three high-mannose oligosaccharides which are possibly recognized by PBP1, the jacalin homolog. It is possible that PBP1 participates in the BGLU23 and ER body-mediated defence systems against herbivores and pathogens (Matsushima *et al.*, 2004). However, PBP1 does not bind to active BGLU23, but acts as a molecular chaperone that helps the correct polymerization of BGLU23 when tissues are damaged. This polymerization of BGLU23 is necessary for its activity (Nagano *et al.*, 2005, 2008).

Unfortunately, the protein interaction data cannot be compared to *in silico* data because no interaction partners for AN4 were found *in silico*. *In silico* interaction data for myrosinase 1 and BGLU23 are available (see discussion BGLU23 above), but AN4 is not known as one of the interaction partners. Of course the binding partners for AN4 retrieved from this pull-down analysis as well as their interaction *in vivo* have to be studied in more detail in the future (Chapter 6).

The three-dimensional models built for Nictaba and AN4 show several resemblances and discrepancies. The canonical β -sandwich core structure of the CBM of GH family 10 enzymes is conserved in both Nictaba and AN4 (Figure 5.8) (Schoupe *et al.*, 2010). Van Holle *et al.* (2017) revealed the three-dimensional model for one of the soybean Nictaba-like lectins and showed that this model also exhibits the canonical β -sandwich core structure. Similar to AN4, the size and shape of the loops connecting the strands of the β -sheet are different from

Nictaba. Furthermore, both Trp-residues important for carbohydrate binding in Nictaba are conserved in the Nictaba-like lectin from soybean (Schouppe *et al.*, 2010; Van Holle *et al.*, 2017a). Considering the sequences of all Nictaba homologs in *A. thaliana* or soybean, both Trp-residues are well conserved, as illustrated with the WebLogos created from all Nictaba domains in *A. thaliana* or soybean. In the WebLogo of the ArathNictabas as well as in the WebLogo of the soybean Nictaba-like lectins, the second Trp-residue is conserved in a higher number of Nictaba homolog sequences than the first one (Eggermont *et al.*, 2017 (Chapter 2); Van Holle *et al.*, 2017a). All these Nictaba homologs with conserved Trp-residues are most likely functional lectins, but their carbohydrate binding specificities are not necessarily conserved as illustrated by the differences in carbohydrate binding specificity between Nictaba from tobacco and an F-box Nictaba homolog from *A. thaliana* (Lannoo *et al.*, 2006; Stefanowicz *et al.*, 2012). Docking experiments performed with chitotriose as a ligand, showed a different localization of the three putative chitotriose-binding sites in Nictaba and AN4. In Nictaba, these three binding sites are localized on a single face of the β -sandwich core structure, while in AN4, both sides of the β -sandwich core structure contain chitotriose-binding sites. Unpublished data from Van Holle S. showed that a Nictaba-like lectin from soybean also contains chitotriose-binding sites at both sides of the β -sandwich core structure, similar to AN4. In Nictaba and AN4, one of these three binding sites showed a higher degree of occupancy, but is situated at the opposite side of the β -sandwich core structure in Nictaba and AN4. Strikingly the chitotriose-binding site with the highest degree of occupancy in a Nictaba-like lectin from soybean is localized at the same side of the β -sandwich core structure as in AN4, namely the concave face (unpublished data Van Holle S.). This is in contrast with Nictaba which contains the chitoriose-binding site with the highest degree of occupancy at the convex face of the β -sandwich core structure. Based on the differences with Nictaba and the similarities with the soybean Nictaba-like lectin, discrepancies in the carbohydrate binding properties between Nictaba and AN4 are hypothesized, while similarities in the carbohydrate binding properties between AN4 and the soybean Nictaba-like lectin are suggested.

Chapter 6

General discussion and future perspectives

Plants are constantly exposed to a plethora of different environmental stresses. Being sessile, they cannot avoid these abiotic and biotic stresses. Therefore plants evolved complicated adaptive and defence mechanisms which allow them to survive in unfavourable conditions (Osakabe *et al.*, 2013). In agriculture, these stresses lead to substantial crop yield losses, but also plants important for the ecosystem are affected by many threats (Gassmann *et al.*, 2016). Because of a rapidly expanding world population, the need for food rises. It is predicted that this global food demand will increase for at least another 40 years because of the continuing population and consumption growth. Furthermore, global warming is an additional threat to food security (Godfray *et al.*, 2010). For all these reasons the elucidation of the underlying defence mechanisms of plants became increasingly important in order to implement this knowledge in the development of more resistant crops.

In the past years, several nucleocytoplasmic lectins from the amaranthin, EUL, GNA, JRL, Nictaba and ricin B lectin families were shown to play a role in signal transduction during plant defence responses to both abiotic and biotic stress (Van Damme *et al.*, 2001, 2004; Wu *et al.*, 2006; Vandenberghe *et al.*, 2010; Xin *et al.*, 2011; Al Atalah *et al.*, 2014; Song *et al.*, 2014; Van Hove *et al.*, 2015). Nictaba from tobacco, the first nucleocytoplasmic lectin discovered within the Nictaba lectin family, is believed to bind *O*-GlcNAc modified core histones upon stress and as such enhances transcription of defence related genes by chromatin remodelling (Schouppe *et al.*, 2011; Delporte *et al.*, 2014).

The major objective of this work was to obtain knowledge about the biological importance of the Nictaba homologs from *Arabidopsis thaliana* (further referred to as ArathNictabas) in the stress responses of *A. thaliana*. Therefore the subcellular localization, the detailed expression pattern under normal growth conditions during plant development as well as after exposure to different stresses, the tolerance of overexpression lines towards stresses and the interaction partners, were investigated. This chapter discusses the most important findings of this PhD thesis and suggests future research experiments necessary to complement this work and reveal the biological importance of the ArathNictabas.

6.1 ArathNictabas are part of the large family of plant lectins and are widespread in angiosperms

A total of 217 putative lectin sequences were defined in the genome of *A. thaliana*. Out of the 217 putative lectins belonging to nine different lectin families, 30 lectin sequences encode homologs of Nictaba from tobacco and thus belong to the Nictaba lectin family (Chapter 2). Very recently a comparative study between four different angiosperms, namely *A. thaliana*, *Glycine max*, *Cucumis sativus* and *Oryza sativa*, was published. This study shows that most plant lectin families, including the Nictaba family, are present in all studied genomes (Van Holle *et al.*, 2017b). Moreover, studies on the Nictaba family revealed the presence of this type of putative lectins in many more angiosperms and even in two mosses, indicating this protein family is widespread throughout the plant kingdom (Delporte *et al.*,

2015; Van Holle *et al.*, 2017a). The widespread distribution of most plant lectin families throughout the whole plant kingdom, indicates the importance of lectins for plants (Van Holle *et al.*, 2017b). In addition, literature data confirm that at least some lectin motifs are active carbohydrate-binding proteins and contribute to plant growth and development (Dang and Van Damme, 2016; Eggermont *et al.*, 2017 (Chapter 2); Van Holle *et al.*, 2017b).

Analysis of the domain architectures for each putative lectin sequence in *A. thaliana* revealed that most sequences contain multiple protein domains. Most of the known protein domains associated with lectin motifs have been reported to be involved in stress signalling, development and defence (Chapter 2). The recent comparative study of Van Holle *et al.* (2017b) also reported a wide range of lectin domain architectures and associations between a lectin motif and one or more domains involved in plant defence, signalling and/or development. This agrees with the experimental data that prove the importance of lectins for plant growth and defence. Although some domain architectures are widespread, others are species-specific. Phylogenetic analyses indicate that domain fusions in a common ancestor are maintained during evolution and give rise to domain architectures that are widespread. In contrast, species-specific retention of a domain architecture or recent events of protein domain fusion or loss in a particular species, most probably accounts for the species-specific domain architectures (Van Holle *et al.*, 2017b). In the genomes studied, non-chimeric Nictabas and F-box Nictabas are widespread, while Nictaba homologs containing a TIR, AIG1-type G, protein kinase or a double Nictaba domain are more species-specific (Delporte *et al.*, 2015; Van Holle *et al.*, 2017a,b).

From the information described in chapter 2 it can be concluded that Arabidopsis plants have a whole range of proteins with lectin motifs at their disposal. The same can be concluded from the comparative study of Van Holle *et al.* (2017b). All these putative lectins are present in different locations in the cell and reside in different plant tissues. Therefore, these putative lectins will most probably exert complementary activities during plant development and survival of the plants under stress conditions (Eggermont *et al.*, 2017 (Chapter 2); Van Holle *et al.*, 2017b).

In the future, it would be interesting to study additional genomes for the presence of lectin motifs from different families, especially from the Nictaba family. Dang (2017) studied 84 plant genomes in their study on distribution of amaranthin-like proteins. The study of additional genomes for the presence of Nictaba homologs would yield more insights into the distribution of this family e.g. among several monocot species (wheat, maize, ...). Moreover, it will be important to prosecute the phylogenetic studies of lectin families across species. These analyses will improve our understanding of the evolutionary history of plant lectins and will continue to discover new protein domain combinations. In principle, each identified putative lectin should be characterized, its interaction with carbohydrate structures should be investigated and the physiological role has to be studied to decipher the importance of all lectins in plants.

6.2 ArathNictabas are low expressed in the nucleus and the cytoplasm

The localization pattern for three non-chimeric ArathNictabas was determined using EGFP fusion proteins (Chapter 3). The results obtained by transient transformation of *N. benthamiana* leaves were confirmed by stable transformation of *A. thaliana* plants (Figure 3.3, Figure 3.4). AN4 and AN5 showed a nucleocytoplasmic localization whereas AN3 only resided in the cytoplasm (Figure 6.1). Cayla *et al.* (2015) also reported a nucleocytoplasmic localization for AN5 using EGFP fusion constructs, which is confirmed by our microscopic analysis. Concerning the subcellular localization of AN3 and AN4, no other experimental data are available, but a cytoplasmic localization is also predicted by the SUBA3 server (Hooper *et al.*, 2014).

The results from the microscopic analyses are in agreement with the absence of a signal peptide in the three ArathNictaba sequences and suggest that translation of these ArathNictabas takes place on free ribosomes in the cytoplasm. Besides the cytoplasm AN4 and AN5 also locate to the nucleus though no classical NLS is found in these ArathNictaba sequences. Lange *et al.* (2007) reported that proteins larger than 40 kDa are too big to diffuse passively into the nuclear compartment, as already shown in 1975 (Paine *et al.*, 1975). In 1998, Görlich (1998) reported that the nuclear pore complex allows passive diffusion from proteins up to 60 kDa. Wang and Brattain (2007) extended this limit, stating that the size of the proteins that diffuse passively can be even larger than 60 kDa. The calculated molecular mass of the EGFP fusion proteins for AN4 and AN5 is 48.5 and 56.3 kDa, respectively. Considering the most recent information about the passive diffusion limit, both proteins could enter the nucleus by passive diffusion through nuclear pore complexes (Wang and Brattain, 2007). Alternatively, AN4 and AN5 possibly contain a non-classical NLS sequence recognized by importin α (Kosugi *et al.*, 2009). Furthermore, additional nuclear import pathways, independent on importin α , have been characterized (Suh and Gumbiner, 2003; Ziemienowicz *et al.*, 2003; Pemberton and Paschal, 2005). Until now, it is not known how the nuclear import of these ArathNictabas works. Similarly, the mechanisms for the nuclear import of Nictaba from tobacco, the two Nictaba-like lectins from soybean and many other nucleocytoplasmic lectins have not yet been unravelled (Al Atalah *et al.*, 2011; Van Hove *et al.*, 2011; Delporte, 2013; Van Holle *et al.*, 2016).

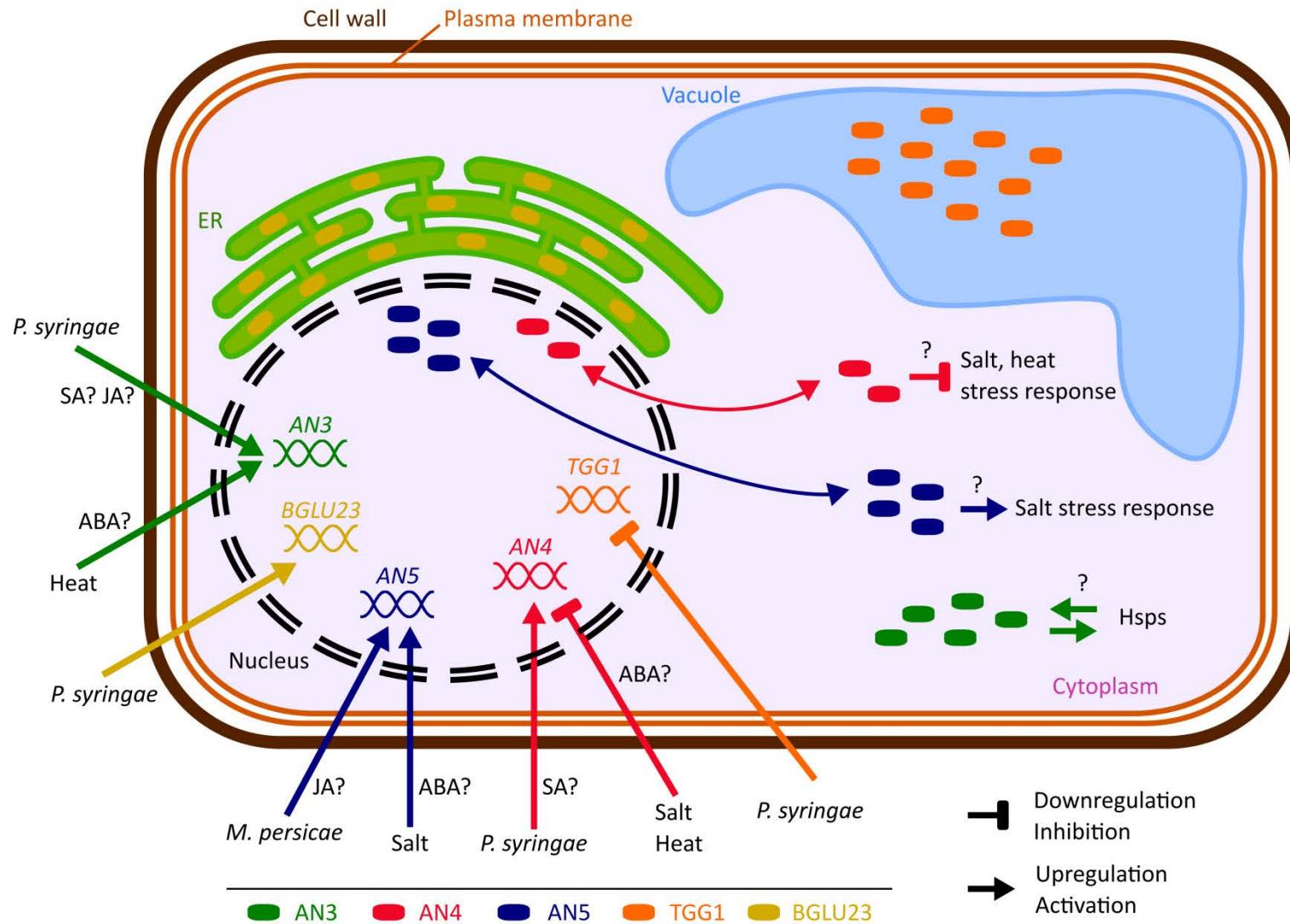


Figure 6.1 Hypothetical model representing the biological importance of the ArathNictabas in *A. thaliana* cells. The cell and organelles are not drawn to scale. Data available from eFP browser suggest that *BGLU23* is upregulated and *TGG1* is downregulated after *Pseudomonas* infection.

Future experiments are necessary to determine whether AN4 and AN5 are actively transported or can passively diffuse from the cytoplasm into the nucleus. The use of compounds inhibiting the active nuclear import would allow to differentiate between active transport and passive diffusion (Soderholm *et al.*, 2011; Liashkovich *et al.*, 2012). Although analysis of the sequences of AN3, AN4 and AN5 revealed no signal peptide or transmembrane region (Chapter 2), a future experiment to exclude localization of AN3, AN4 and AN5 to the plasma membrane and cell wall is advisable since our microscopic analysis does not allow to distinguish between the cytoplasm, the plasma membrane and the cell wall. Analysis of the EGFP fluorescence in protoplasts from the stably transformed *A. thaliana* ArathNictaba EGFP lines and treatment of these protoplasts with high concentrations of mannitol or salt can confirm the localization of the three ArathNictabas to the cytoplasm. A similar experiment with stably transformed BY-2 cells allowed to confirm the cytoplasmic localization of Nictaba from tobacco (Delporte, 2013).

Using qRT-PCR, the expression level of the *ArathNictabas* was investigated in different tissues from *A. thaliana* during development. The three *ArathNictaba* genes are expressed in every tissue during all developmental stages tested. It should be mentioned that lectin expression during the development of the plant is generally low, especially for AN3 and AN4 (Chapter 3). Overall the expression level of AN5 is higher than the expression levels of AN3 and AN4 throughout plant development (Figure 3.7). However, specific stress factors can enhance the transcript levels for each lectin. The low expression observed for the *Arathnictabas* is unlike the expression patterns observed for *Nictaba* from tobacco. The tobacco lectin could not be detected at all in plants grown under non-stress conditions (Chen *et al.*, 2002; Lannoo *et al.*, 2007a). Similar to AN3, AN4 and AN5, *F-box Nictaba* from *A. thaliana* and the *Nictaba*-like lectins from soybean were present in plant cells at low levels when grown under optimal growth conditions (no stress) (Stefanowicz *et al.*, 2016; Van Holle, 2016; Van Holle *et al.*, 2016).

Future experiments should investigate different phenotypic traits of Arabidopsis using *ArathNictaba* mutants. The available mutants from NASC have been tested in this PhD thesis, but did not contain a T-DNA insertion in the ArathNictaba genes. Therefore mutants have to be generated through mutation of specific amino acid residues. Alternatively transgenic lines with reduced lectin expression can be generated using e.g. the RNAi technology (Younis *et al.*, 2014). A new, now commonly used technique, to make knock out mutants is the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) system (Bortesi and Fischer, 2015). Furthermore the classical promoter-GUS experiments could confirm the expression of the *ArathNictabas* in different tissues during the development of the plant, but could also yield additional information related to lectin expression in specific cell types (Jefferson *et al.*, 1987).

6.3 The expression of *ArathNictabas* is stress-inducible

Plant stress is generally classified into abiotic and biotic stresses. Each stress, abiotic or biotic, evokes a complex cellular and molecular response in the plant to prevent damage and ensure survival of the plant. Hormones act as signalling molecules in these plant stress responses (Atkinson and Urwin, 2012; Huber and Bauerle, 2016). In chapter 3 WT *Arabidopsis* plants were subjected to different hormone, abiotic or biotic stress treatments and the expression level of three non-chimeric *ArathNictabas* was quantified by qRT-PCR.

When the stress-inducible expression pattern for the different *Nictaba* homologs is compared, it can be concluded that expression patterns for the *ArathNictabas* are specific and vary for the different abiotic or biotic stress treatments performed (Chapter 3) (Figure 6.1).

6.3.1 Abiotic stress

In chapter 3 *Arabidopsis* seedlings were subjected to salt and heat stress, two major abiotic stresses. High salinity stress appears primarily as osmotic stress and as such results in the disruption of homeostasis and ion distribution in the cell (Wang *et al.*, 2003). Heat stress results in the global inhibition of translation and secondly plants have to cope with osmotic and oxidative stress (Wang *et al.*, 2003; Qu *et al.*, 2013; Echevarría-Zomeño *et al.*, 2016). The phytohormone ABA is a key regulator in the plant stress response to osmotic stress (Goldack *et al.*, 2014). Moreover, ABA also plays an important role in heat stress (Vishwakarma *et al.*, 2017).

The expression level of *AN3* is generally highly upregulated after heat stress, up to six fold, whereas ABA treatment yields a twofold upregulation of the expression level of *AN3* after 5 and 10 hours (Chapter 3). Taking into account that the expression level of *AN3* after salt treatment is not changed, *AN3* most probably does not play an essential role in the ABA-dependent salt stress response. Not surprisingly the promoter of *AN3* contains one HSE (Chapter 3). Heat responsive genes with HSEs are controlled by Hsf TFs. These heat responsive genes are mostly protein kinases, TFs, Hsps or catalases (Wang *et al.*, 2003; Qu *et al.*, 2013; Echevarría-Zomeño *et al.*, 2016). Taking into account that *AN3* has a lectin domain but does not contain any other known protein domain, it is unlikely that *AN3* is a protein kinase, TF, Hsp or catalase. It is more likely that *AN3* has a similar function as Hsps, chaperones which play a role in preventing protein unfolding and misfolding during heat stress or that *AN3* assists the Hsps in recognizing their targets. Alternatively, *AN3* can also play a role downstream of the main heat responsive genes: protein kinases, TFs, Hsps and catalases (Figure 6.1).

Future experiments will be necessary to unravel the role of *AN3* in the heat stress response. It would be interesting to analyse the expression pattern of *AN3* after heat stress in an ABA deficient mutant e.g. the *aba2* mutant, where the ABA2 enzyme, one of the enzymes

necessary for ABA biosynthesis, is mutated (Léon-Kloosterziel *et al.*, 1996). This experiment will give more information on the role of ABA and AN3 as part of heat stress signalling. Since three independent overexpression lines for AN3 are available (Chapter 4), it could be envisaged to test the thermotolerance of these overexpression lines. Additionally *Arabidopsis* mutants for AN3, if first generated, can be part of this experiment. Of course the survival rate is one of the factors that has to be studied in these thermotolerance experiments, but next to that root length, ROS levels, proline and chlorophyll content can be measured (Krčková *et al.*, 2015; Wu *et al.*, 2015; Suzuki *et al.*, 2016).

Overall the expression level of AN4 is downregulated after ABA treatment, salt and heat stress (Chapter 3). Downregulation of transcript levels during stress responses is more difficult to interpret, but possibly AN4 has a negative effect on the salt and heat stress response of the plant (Figure 6.1). This negative effect can be due to the repression of specific steps in the stress signalling pathway. If this is really the case, overexpression of AN4 could negatively influence the tolerance of these overexpression lines towards salt stress. In chapter 4, the stress tolerance of transgenic lines was tested in a germination assay, and indeed revealed a better germination of the WT plants compared to the overexpression lines for AN4. However, this result has to be interpreted with care because seeds from all overexpression lines for AN5 also showed a significantly lower germination than the WT. Nevertheless the expression level of AN5 is upregulated after salt stress, as discussed below.

Future experiments will have to unravel if AN4 really has a negative impact on the stress responses provoked by salt and heat stress. More experiments with the overexpression lines could give more insight in these signalling processes. To avoid problems with differences in germination between different seed batches, seedlings could first be germinated on medium without salt. After germination, the seedlings could then be transferred to a medium containing salt and the performance of the overexpression lines and the WT plants could be compared. Furthermore, it would be interesting to include AN4 mutants in these stress tolerance experiments, similar to what has been discussed above for AN3.

The expression level of AN5 shows a 1.7 fold significant upregulation after salt treatment for 10 hours and thus not surprisingly a 2.3 fold upregulation was observed after treatment with ABA for 10 hours (Chapter 3). This can be an indication that AN5 plays a role in the ABA-dependent pathway of the salt stress response (Figure 6.1). The promoter sequence of AN5 also contains a drought responsive MYB binding site, several TFs of the MYB family can induce ABA-responsive genes (Chapter 3) (Wang *et al.*, 2003; Bhargava and Sawant, 2013; Roychoudhury *et al.*, 2013). As already mentioned above, the tolerance of three independent AN5 overexpression lines was tested in a germination assay in the presence of salt and revealed a lower germination compared to the WT seeds. This is not what is expected if AN5 would play a role in the ABA-dependent response upon salt stress.

In future experiments it would be interesting to quantify the expression levels of AN5 in *aba2* mutants after salt stress (Léon-Kloosterziel *et al.*, 1996). If these expression levels are

still upregulated, the role of AN5 in the salt response is not purely ABA-dependent. In addition, more experiments are needed to study the role of AN5 in the salt stress response. Surely mutants for AN5 would be an added value for the salt tolerance experiments.

6.3.2 Biotic stress

Five-week-old *Arabidopsis* plants were subjected to *P. syringae*, *B. cinerea* infection and *M. persicae* infestation (Chapter 3). *P. syringae* is a hemibiotroph that belongs to the gram-negative plant-pathogenic bacteria (Surico, 2013; Büttner, 2016). *B. cinerea* is a necrotrophic fungus and produces diverse phytotoxic compounds and cell-wall degrading enzymes to induce cell necrosis and as such leakage of nutrients (Mengiste, 2012). *M. persicae* is an insect belonging to the class of pierce-sucking insects and uses its stylet to feed from the phloem of the plant. It is a generalist which means it can feed on large variety of plants belonging to different families (Louis and Shah, 2013). Two phytohormones play a major role in the defence against biotic stresses. *P. syringae* infection is activating the SA-dependent plant defence pathway while *B. cinerea* infection and *M. persicae* infestation are activating the JA-dependent plant defence pathway (Katagiri *et al.*, 2002; Howe and Jander, 2008; Mengiste, 2012; Pieterse *et al.*, 2012).

The expression level of AN3 shows an early 2 - 2.5 fold upregulation after *P. syringae* infection (after 3 days) and a twofold downregulation after 3 days of *M. persicae* infestation. The upregulation upon *P. syringae* infection correlates with the 1.5 - 3.5 fold upregulation of AN3 after SA treatment for 1 - 24 hours, but does not correlate with the 2 - 4 fold upregulation of the expression of AN3 after 1 - 24 hours MeJA treatment (Chapter 3) (Figure 6.1). For both hormones the expression is overall significantly upregulated, which is confusing because the SA- and JA-dependent defence pathways normally work antagonistically (Smith *et al.*, 2009; Vos *et al.*, 2013). That is why *P. syringae* even produces coronatine, a compound which is structurally similar to JA, and was shown to suppress the SA-mediated defence of the plant (Katagiri *et al.*, 2002; Jones and Dangl, 2006). Hormone cross-talk is very complex which is illustrated by the fact that also neutral and synergistic interactions between SA and JA have been reported. Timing, sequence of initiation and the relative concentration of each hormone play a role in the outcome of the SA-JA cross-talk (Vos *et al.*, 2013). The overexpression lines of AN3 did not reveal a reduced susceptibility to *P. syringae* infections compared to the WT plants. In the two biological replicates performed with 35S::AN3 plants, the *Arabidopsis* leaves of the WT plants showed less disease symptoms after infection with *P. syringae* compared to the WT plants in the infection experiments with the overexpression lines for AN4 and AN5 (Chapter 4). Probably a stronger bacterial infection is needed to show whether the overexpression of AN3 can yield a reduced susceptibility of the plants towards *P. syringae* infections.

In the future, the role of AN3 in the SA- and JA-dependent defence pathways has to be investigated in more detail. First of all the involvement of AN3 in these defence pathways has to be confirmed by studying the expression of AN3 in mutants of the SA and JA pathway. Furthermore, the role of AN3 in the defence against *P. syringae* has to be resolved. This can be achieved by using SA deficient mutants in the *P. syringae* infection experiments and determining the AN3 expression level after *P. syringae* infection.

The expression level of AN4 is upregulated two times at 5 and 7 days after *P. syringae* infection, but in general the expression level of AN4 is not changed after SA treatment, indicating that AN4 might play a role in the plant response to *P. syringae*, independent of SA (Chapter 3) (Figure 6.1). It is unclear whether AN4 can act in the defence of Arabidopsis against *P. syringae*, independent of SA, since only one of the three overexpression lines for AN4 shows a reduced susceptibility than WT plants towards *P. syringae* infection (Chapter 4). The better performance of transgenic line AN4_B1 can also be an off-target effect e.g. related to the place where the T-DNA insertion in the genome took place. The significant downregulation of AN4 in the MeJA treated Arabidopsis seedlings is small and does not correlate with the expression of AN4 upon *M. persicae* infestation (Chapter 3).

Further research is necessary to unravel the effect of *P. syringae* infection on the expression of AN4. It will have to be investigated if the expression level of AN4 is still upregulated after *P. syringae* infection in SA deficient mutants. If so, it can be concluded that this response is really SA-independent. Furthermore, *P. syringae* infection experiments with AN4 mutants would be interesting to see whether these mutants are more susceptible to *P. syringae* infection than WT plants.

The expression level of AN5 shows an almost twofold upregulation after 3 days of *M. persicae* infestation. This is in agreement with the small upregulation of AN5 expression observed after MeJA treatment (Chapter 3). Possibly AN5 plays a role in the JA-dependent defence against aphids (Figure 6.1) (Louis and Shah, 2013). This role is in agreement with the repression in phloem-feeding activities of *M. persicae* as a result of overexpression of AN5 in Arabidopsis (Zhang *et al.*, 2011). Moreover Beneteau *et al.* (2010) showed that recombinant AN5 at mid-range concentrations, affects weight gain of *M. persicae* nymphs. The SA signaling pathway is also known to be activated during aphid infestation, but this pathway facilitates aphid infestation (Louis and Shah, 2013). Transcriptome analysis of Arabidopsis subjected to another phloem-feeding insect namely *Bemisia tabaci*, the silverleaf whitefly, revealed differences in the plant responses after silverleaf whitefly and *M. persicae* infestation. Indeed, also for AN5 the response is different (opposite), since the expression of AN5 showed a twofold downregulation after *B. tabaci* infestation (Kempema *et al.*, 2006). All overexpression lines for AN5 are more tolerant (not always significant) to *P. syringae* infections compared to WT plants. Since all three independent AN5 overexpression lines show similar results and are more tolerant, this result is probably related to the

overexpression of *AN5* (Chapter 4). This result is surprising because the expression level of *AN5* is not changed after *P. syringae* infection.

In order to elucidate the role of *AN5* in the plant defence against *M. persicae* in the future, the survival, offspring and development of the aphids have to be investigated on *AN5* overexpression lines and mutants. Moreover JA deficient mutants can be infested with *M. persicae* to test *AN5* expression levels and unravel the role of *AN5* in the aphid stress response.

6.3.3 Cross-talk between abiotic and biotic stress

The phytohormones known to be important in abiotic and biotic stress responses often interact with each other. Moreover these hormones can play a role in both abiotic and biotic stress. JA, SA and ET play a role in abiotic stresses (Miura and Tada, 2014; Dar *et al.*, 2015; Kazan, 2015; Valenzuela *et al.*, 2016), but ABA is also important in the defence against biotic stresses (Tsuda and Katagiri, 2010; Pieterse *et al.*, 2012; Vos *et al.*, 2013). In addition, several growth hormones were also reported to play a role in abiotic and biotic stress, making the hormone signalling in plants subjected to stress even more complex (Tsuda and Katagiri, 2010; Bhargava and Sawant, 2013; Kazan, 2015; Couto and Zipfel, 2016). Next to these phytohormones, two other signals play a role in the cross-talk between abiotic and biotic stresses namely Ca^{2+} and ROS (Sakamoto *et al.*, 2008; Tsuda and Katagiri, 2010; Bhargava and Sawant, 2013; Muthamilarasan and Prasad, 2013; Couto and Zipfel, 2016; Vishwakarma *et al.*, 2017).

In nature, plants are often simultaneously exposed to multiple stresses, abiotic as well as biotic stresses often occur together. Until now most studies focused on the effect(s) of separate stresses, but our understanding of the stress signalling pathways under combinations of abiotic and biotic stresses is still rather poor. The presence of an abiotic stress can have a positive or a negative effect on the susceptibility to a biotic agent, and vice versa. Moreover the response of the plant to a combination of stresses can be totally different from the response to each of the individual stresses. Therefore it will be important to study the expression of *ArathNictabas* in plants subjected to a combination of stresses. The study of these combination(s) of stresses is fundamental to develop broad-spectrum stress tolerant crop plants (Atkinson and Urwin, 2012; Rasmussen *et al.*, 2013).

6.4 AN4 interacts with two plant defence-involved enzymes

The genome of *A. thaliana* is fully sequenced and annotated. For many proteins, interaction partners are predicted and/or experimentally proven, revealing part of the interactome of *A. thaliana*. Until now, no interaction partners are reported for *AN4*. Pull-down experiments using *AN4*-HIS coupled to Ni agarose beads as a bait protein and an Arabidopsis plant extract as prey proteins, followed by MS analysis revealed two possible interaction partners for *AN4*, namely TGG1 and BGLU23 (Figure 6.1). At first sight, there is no obvious link between *AN4*

and the two plant defence-involved enzymes TGG1 and BGLU23. However, interesting observations were made when the localization, expression pattern and levels were compared in detail.

Protein localization and expression under normal growth conditions can be compared at three levels, in particular at subcellular level, according to cell type and at tissue level. AN4 resides in the nucleus and the cytoplasm of the cell and its expression under normal growth conditions is significantly higher in the roots compared to 6-day-old seedlings (Chapter 3). Jakoby *et al.* (2008) identified AN4 as one of the 5 % most highly expressed genes in mature *Arabidopsis* trichomes. Moreover, they reported a strong enrichment of genes involved in root atrichoblast differentiation in the trichome transcriptome (Jakoby *et al.*, 2008). The sequence encoding TGG1 contains a signal peptide and the protein localizes to the vacuole (Thangstad *et al.*, 2004). Xue *et al.* (1995) reported the expression of TGG1 in the leaves, stems and floral organs. At cell type level, TGG1 was especially found in the guard cells and in phloem myrosin cells. Myrosin cells or idioblasts are individual, specific cells that differ greatly from their neighbours in size, structure and content. The morphology of these specific cells can be different in different organs, tissues and developmental stages (Thangstad *et al.*, 2004). BGLU23, like TGG1, is synthesized with a signal peptide, but also contains an ER-retention signal (KDEL sequence). This β -O-glucosidase, also called PYK10 is specifically localized in ER bodies and is root specific (Nitz *et al.*, 2001; Matsushima *et al.*, 2003; Ahn *et al.*, 2010). BGLU23 was also found in the *Arabidopsis* plasmodesmal proteome, but probably represents a cytoplasmic contaminant in this fraction (Fernandez-Calvino *et al.*, 2011).

A striking similarity between AN4 and BGLU23 at tissue level is their expression in the roots of *Arabidopsis*. However, according to microarray data (eFP browser) the expression level of *BGLU23* in the roots is much higher than the expression level of *AN4* under normal growth conditions (Winter *et al.*, 2007). Furthermore, it is interesting to note that *BGLU23* was also identified as one of the 5 % most highly expressed genes in mature *Arabidopsis* trichomes (Jakoby *et al.*, 2008). Although trichomes are present on the leaves, it is shown that trichome development in leaves and atrichoblast development in roots share a network of TFs (Pesch and Hülkamp, 2004; Schellmann *et al.*, 2007). However, it is not known if trichomes and atrichoblasts really share common patterns of gene expression (Jakoby *et al.*, 2008). Jakoby *et al.* (2008) reported high activities of genes involved in the glucosinolate pathways in trichomes, indicating the role of trichomes in plant defence. Possibly AN4 and BGLU23 can interact in the roots or trichomes when the plant is subjected to stress. To do so, at least one of the two proteins will have to change its subcellular localization as a result of the stress. Changing subcellular localization upon stress was reported before (García *et al.*, 2010; Moore *et al.*, 2011; Sabol *et al.*, 2017). Alternatively AN4 and BGLU23 can interact as a result of the cell damage provoked by e.g. pathogens.

Another similarity at tissue level is seen between TGG1 and AN4. TGG1 was especially abundant in the phloem myrosin cells whereas AN4 was reported as a PP2 of Arabidopsis (Dinant *et al.*, 2003; Thangstad *et al.*, 2004). PP2 proteins are most abundant in the phloem sap. Though the sequence of AN4 shows high similarity to the PP2 domain, a protein domain that is conserved among many species in the plant kingdom, there is no evidence for the presence of AN4 in the phloem, except for the microarray data from eFP browser (Dinant *et al.*, 2003; Winter *et al.*, 2007).

As already mentioned in chapter 3, the expression level of AN4 is low under normal growth conditions and the expression of AN4 stays low or is even downregulated in all stress treatments investigated, except for the *P. syringae* infection. *P. syringae* infection of 5-week-old rosette leaves resulted in approximately two times upregulation of AN4 transcript levels at the latest timepoints (after 5 and 7 days). According to the microarray data of the eFP browser, BGLU23 shows a sixfold upregulation 24 hours after *P. syringae* infection, but TGG1 shows a small downregulation (Winter *et al.*, 2007). It has to be mentioned that in our data AN4 also first shows a twofold significant downregulation 24 hours after *P. syringae* infection. Although TGG1 and AN4 show a similar regulation after 24 hours *P. syringae* infection, their absolute expression levels are very different, TGG1 is present in much higher amounts in infected Arabidopsis leaves than AN4 (Figure 6.1). Furthermore, it remains unclear if the expression of BGLU23 and AN4, and their interaction in the trichomes is part of the defence response against Pseudomonas.

Since the pull-down analysis is an *in vitro* analysis and the plant extract is a mixture of proteins derived from different cell compartments, it is not certain that the interaction partners retrieved from the pull down assay represent true *in vivo* interaction partners for AN4. Additionally, TGG1 and BGLU23 are two proteins that are quite abundant in Arabidopsis seedlings, whereas AN4 is expressed at a much lower level. Future experiments are thus necessary to confirm interaction *in vivo*. One technique to confirm interactions *in vivo* is bimolecular fluorescence complementation (BiFC) (Kerppola, 2008). Since the localization of AN4, TGG1 and BGLU23 is experimentally confirmed and they all have a different subcellular localization, BiFC experiments under normal growth conditions will be probably yield a negative result. It will be important to subject the plants to the right stress condition in order to check whether AN4 can interact with TGG1 and/or BGLU23 under these stress conditions.

Once the interaction of AN4 with TGG1 and/or BGLU23 is confirmed *in vivo*, it would be interesting to study mutants of TGG1 and/or BGLU23. The expression levels of AN4 could be compared between WT plants and these mutant lines. Once it is known in which stress response the interaction of AN4 with TGG1 and/or BGLU23 is involved, the tolerance of these mutants and of double mutants of AN4 and TGG1 or AN4 and BGLU23 can be investigated. The enzyme activities of TGG1 and BGLU23 could also be tested in AN4

mutants in order to investigate whether AN4 possibly plays a role as a co-factor e.g. in the recognition of glucosinolates.

Until now, no carbohydrate-binding activity could be demonstrated for AN4, but the molecular modelling of the AN4 sequence and docking with chitotriose indicated three different putative chitotriose-binding sites on the molecular surface of AN4 (Chapter 5). The interaction of AN4 with carbohydrate structures has to be confirmed experimentally.

In the future, it is worthwhile to optimize the conditions to produce more recombinant AN4. The recombinant protein can be used for a glycan array analysis. Knowledge of the carbohydrate binding specificity is important in order to unravel the biological importance of AN4. Indeed, the identified interaction partners TGG1 and BGLU23 are synthesized following the secretory pathway and thus can be glycosylated. Sequence analysis for TGG1 and BGLU23 identified eight and three putative *N*-glycosylation sites in the sequences of TGG1 and BGLU23, respectively. Whether TGG1 and BGLU23 are glycosylated *in vivo* is not known. TGG1 and BGLU23 proteins overexpressed in *Arabidopsis* could be analyzed with MS in the future in order to identify if these proteins are glycosylated and if so, which glycan structures are present. PNGase F could be used to cleave off the carbohydrates of TGG1 and BGLU23. Subsequently, it can be tested if deglycosylated TGG1 and BGLU23 are still retained on a column with immobilized AN4. This result will allow to determine whether the interaction between AN4 and TGG1/BGLU23 is a protein-protein interaction or relies on a protein-carbohydrate recognition.

Nictaba from tobacco is believed to interact with *O*-GlcNAc modified core histones. It is hypothesized that this interaction allows Nictaba to remodel the chromatin and alter the transcription of certain stress-responsive genes. It remains to be investigated which genes are regulated by this Nictaba-histone interaction. The original goal of this PhD thesis was to study if the *Arath*Nictabas display a similar role in the plant stress response as Nictaba from tobacco. However, this PhD research revealed that the genome of *A. thaliana* contains a much larger set of Nictaba homologues than the tobacco genome. Furthermore the study of the non-chimeric *Arath*Nictabas also revealed important differences in the regulation of gene expression and the protein characteristics for the Nictaba homologs in both species.

Summary

Plants are constantly exposed to a wide range of environmental stresses, but evolved complicated adaptive and defence mechanisms which allow them to survive in unfavourable conditions. Of course the plant needs energy to activate this complicated defence system and loses energy that would have been used for plant growth. Reduced plant growth, especially of the vegetative tissues, causes crop yield losses in agriculture which can be prevented by using stress-tolerant crops. *Arabidopsis thaliana*, the most widely-used model plant, is a valuable system to study the underlying mechanisms of the complicated plant defence system. Afterwards this knowledge can be extrapolated to crops and used to develop modified crops with an improved tolerance to unfavourable environmental conditions. Lectins or carbohydrate-binding proteins constitute an important part of this plant defence system. Lectins are widespread in the plant kingdom and can be grouped into vacuolar lectins with a constitutive expression and nucleocytoplasmic lectins with a stress-inducible expression. In the past years, lectin research has focused on the stress-inducible nucleocytoplasmic lectins. The *Nicotiana tabacum* agglutinin, abbreviated as Nictaba, served as a model for one family of stress related lectins. This PhD work focused on several Nictaba homologs from *Arabidopsis thaliana*, further referred to as the ArathNictabas. The localization, expression pattern and interaction partners for these ArathNictabas have been investigated to get insight into the biological importance of these proteins for the stress responses of *A. thaliana*.

Chapter 1 presents a literature overview of the most important plant stresses and the responses which allow plants to survive in unfavourable circumstances. In addition, plant lectins are shortly described with the focus on the lectins belonging to the Nictaba family.

The genome of *Arabidopsis* is fully sequenced and annotated. In **chapter 2** a genome wide screening for lectin motifs was performed to identify all putative lectin sequences in the *Arabidopsis* genome. In total 217 putative lectin genes were retrieved belonging to nine out of twelve plant lectin families identified today. The domain architectures of all putative lectins were determined and discussed together with available information from literature concerning their biological functions. Domain architecture analysis revealed that most of these lectin gene sequences are linked to other domains and many of these domains are known to play a role in stress signalling and defence. Phylogenetic analysis of the Nictaba-related lectins, the jacalin-related lectins and the LysM-related lectins showed that lectin sequences that share the same domain architecture often evolved together. Furthermore, sequence analysis of these putative lectin sequences revealed that the amino acids responsible for carbohydrate binding activity are largely conserved.

In **chapter 3** a thorough expression analysis was performed for three non-chimeric *ArathNictabas*, further referred to as *AN3*, *AN4* and *AN5*, in order to determine their subcellular localization and the expression pattern under normal growth conditions as well as after different (a)biotic stress treatments. To study the subcellular localization EGFP fusion constructs for the *ArathNictaba* sequences were created and transiently transformed

in *Nicotiana benthamiana* leaves or stably transformed in *A. thaliana* plants. Confocal microscopy revealed fluorescence for AN4 and AN5 in the nucleus and the cytoplasm while fluorescence for AN3 was only detected in the cytoplasm. Transcript levels for the *ArathNictabas* were determined in different tissues during normal development and in plants subjected to various abiotic and biotic stress treatments. qRT-PCR analysis revealed a low expression for all three *ArathNictabas* in different tissues throughout plant development. Stress application altered the expression levels for the *ArathNictabas*, but all three *ArathNictabas* showed a different expression pattern. Taken together, our data suggest that the *ArathNictabas* represent stress regulated proteins with a possible role in plant stress responses.

In **chapter 4** *ArathNictaba* overexpression lines were generated in Arabidopsis and subjected to salt stress and bacterial infection in order to elucidate whether the *ArathNictaba* genes can help the plant to cope with these stress situations. Germination experiments with *ArathNictaba* overexpression lines under salt stress conditions revealed no better tolerance of the transgenic plants towards mild and high salt stress compared to wildtype plants. *P. syringae* infection experiments with the *ArathNictaba* overexpression lines revealed that several transgenic lines for AN4 and AN5 showed a significantly lower level of leaf damage, a lower percentage of cell death and reduced levels of Pseudomonas biomass. From these results it can be concluded that two of the three *ArathNictabas* possibly play a role in the defence of the plant against *P. syringae*.

Chapter 5 presents the recombinant production of the *ArathNictaba* proteins. The recombinant AN4 protein produced in *Escherichia coli* was used to investigate lectin activity. It was not possible to show agglutination activity for the purified AN4 in assays with rabbit erythrocytes. In addition, no carbohydrate binding activity was detected with immobilized carbohydrates and glycoproteins. A pull-down assay followed by MS analysis was performed to identify interaction partners for AN4. MS analysis of the proteins retrieved from the pull-down analysis identified two possible interaction partners, namely myrosinase 1 (TGG1) and β -glucosidase 23 (BGLU23/PYK10). Similar to AN4, these two interaction partners play a role in the plant defence.

Chapter 6 discusses the most important findings of this PhD thesis and suggests future research experiments necessary to complement this work and reveal the biological importance of the *ArathNictabas*. To conclude, the data in this PhD thesis show that *ArathNictabas* are involved in the plant defence response directed towards different abiotic and biotic stresses. This research contributed to the elucidation of underlying mechanisms in plant stress responses and on long term can contribute to the development of more stress-resistant plants.

Samenvatting

Planten worden voortdurend blootgesteld aan verschillende stresssituaties in hun omgeving, maar hebben defensiemechanismen ontwikkeld die ervoor zorgen dat ze kunnen overleven. Planten hebben energie nodig om deze defensiemechanismen te activeren en verliezen hierbij energie die normaal gebruikt wordt voor de groei van de plant. Deze verminderde groei van de vegetatieve plant maar ook van de vruchten kan leiden tot opbrengstverliezen in de landbouw. Een mogelijke oplossing hiervoor is het gebruik van stresstolerante gewassen. *Arabidopsis thaliana* is wereldwijd het meest bestudeerde modelorganisme binnen de planten en is zeer waardevol om de onderliggende mechanismen van het plant defensiesysteem te bestuderen. Deze kennis kan vervolgens geëxtrapoleerd worden naar landbouwgewassen en zo gebruikt worden om gemodificeerde gewassen te ontwikkelen met een verhoogde tolerantie tegen stress. Lectines of suikerbindende eiwitten maken een belangrijk deel uit van het plant defensiesysteem. Lectines zijn wijdverspreid in het plantenrijk en kunnen gegroepeerd worden in vacuolaire lectines met een constitutieve expressie en nucleocytoplasmatische lectines met een stress-induceerbare expressie. Afgelopen jaren focuste het onderzoek naar lectines zich op de stress-induceerbare nucleocytoplasmatische lectines. Het *Nicotiana tabacum* agglutinine, afgekort als Nictaba, is het model binnen één familie van deze stress-gerelateerde lectines. Deze doctoraatsthesis onderzoekt een aantal Nictaba homologen uit *Arabidopsis thaliana*, verder vernoemd als ArathNictaba's. De lokalisatie, het expressiepatroon en de interactiepartners voor deze ArathNictaba's werden onderzocht om inzicht te krijgen in hun rol in de reactie van *A. thaliana* op stresssituaties.

Hoofdstuk 1 bevat een literatuuroverzicht van de belangrijkste soorten stress bij planten en de mechanismen die ervoor zorgen dat planten kunnen overleven in deze stresssituaties. Daarenboven worden plantenlectines kort beschreven, waarbij de nadruk gelegd wordt op lectines van de Nictaba familie.

Het genoom van *Arabidopsis* is volledig gesequeneerd en geannoteerd. In **hoofdstuk 2** werd een screening uitgevoerd van het volledige *Arabidopsis* genoom om alle lectinesequenties te identificeren. In het totaal werden 217 mogelijke lectinegenen gevonden welke tot negen van de twaalf op heden gekende planten lectinefamilies behoren. De domeinarchitecturen van deze lectinesequenties werden bepaald en de beschikbare literatuurgegevens met betrekking tot hun biologische functie werd besproken. De analyse van de domeinarchitecturen toonde aan dat de meeste van deze lectinesequenties, naast het lectinedomein ook één of meerdere andere domeinen bevatten. Vele van deze domeinen spelen een rol in stresssignalering en defensiemechanismen. Een fylogenetische analyse van de Nictaba-gerelateerde lectines, de jacaline-gerelateerde lectines en de LysM-gerelateerde lectines toonde aan dat lectinesequenties met dezelfde domeinarchitectuur vaak samen geëvolueerd zijn. Bovendien toonde een sequentie-analyse van deze lectinesequenties aan dat de aminozuren die verantwoordelijk zijn voor de suikerbindende activiteit sterk geconserveerd zijn.

In **hoofdstuk 3** werd een expressie-analyse uitgevoerd voor drie niet-chimere *ArathNictaba's* die verder vernoemd worden als *AN3*, *AN4* en *AN5*. De subcellulaire lokalisatie en het expressiepatroon onder normale groei-omstandigheden, alsook na verschillende (a)biotische stressbehandelingen werd onderzocht. EGFP fusieconstructen voor de *ArathNictaba* sequenties werden gemaakt en transiënt getransformeerd in de bladeren van *Nicotiana benthamiana* of stabiel getransformeerd in *A. thaliana* planten. Met behulp van confocale microscopie kon er fluorescentie voor *AN4* en *AN5* aangetoond worden in de nucleus en het cytoplasma van de plantencel terwijl de fluorescentie voor *AN3* enkel in het cytoplasma aanwezig was. De expressie van de *ArathNictaba's* werd bepaald in verschillende weefsels gedurende de normale ontwikkeling van de plant, alsook in planten die verschillende abiotische en biotische stressbehandelingen ondergingen. qRT-PCR analyses toonden een lage expressie voor de drie *ArathNictaba's* in de verschillende weefsels gedurende de ontwikkeling van de plant. De stressbehandelingen beïnvloedden de expressie van de *ArathNictaba's*, maar deze veranderingen in expressie waren verschillend voor de drie *ArathNictaba's*. Onze data geven dus aan dat *ArathNictaba's* stress-gereguleerde eiwitten zijn met een mogelijke rol in de reacties van planten op stress.

In **hoofdstuk 4** werden *ArathNictaba* overexpressielijnen gemaakt in Arabidopsis. Deze planten werden blootgesteld aan zoutstress en bacteriële infectie om na te gaan of de *ArathNictaba* genen de plant kunnen helpen om beter met deze stresssituaties om te gaan. De kieming van de *ArathNictaba* overexpressielijnen werd bestudeerd tijdens zoutstress en toonde geen betere tolerantie dan wildtype planten tijdens milde en hoge zoutstress. In infectie-experimenten met *Pseudomonas syringae* werd er wel een significant verlaagde vergeling van de bladeren, een lager celdood percentage en een lagere *Pseudomonas* biomassa aangetoond voor verschillende overexpressielijnen van *AN4* en *AN5*. Uit deze resultaten kan geconcludeerd worden dat twee van de drie *ArathNictaba's* mogelijk een rol spelen in de verdediging van de plant tegen *P. syringae*.

De productie van recombinante *ArathNictaba* eiwitten wordt beschreven in **hoofdstuk 5**. Het recombinant aangemaakt *AN4* eiwit werd geproduceerd in *Escherichia coli* en gebruikt om lectine-activiteit na te gaan. Er kon geen agglutinatie-activiteit aangetoond worden voor het opgezuiverde *AN4* eiwit in de agglutinatie-testen met rode bloedcellen van konijnen. Daarnaast kon er ook geen suikerbinding vastgesteld worden aan geïmmobiliseerde suikers en glycoproteïnen. Interactiepartners voor *AN4* werden geïdentificeerd in een pull-down experiment gevolgd door een MS analyse. In deze MS analyse werden twee mogelijke interactiepartners voor *AN4* geïdentificeerd, namelijk myrosinase 1 (TGG1) en β -glucosidase 23 (BGLU23/PYK10). Deze interactiepartners spelen, net zoals *AN4*, een rol in het plant defensiemechanisme.

In **hoofdstuk 6** worden de belangrijkste resultaten van deze doctoraatsthesis beschreven en worden experimenten voorgesteld die in de toekomst nodig zullen zijn om de biologische functie van de *ArathNictaba's* volledig te ontrafelen. We kunnen besluiten dat de resultaten

van deze doctoraatsthesis aantonen dat Nictaba homologen uit Arabidopsis een rol spelen in defensiemechanismen van de plant tijdens verschillende abiotische en biotische stresssituaties. Ons onderzoek heeft bijgedragen aan de opheldering van de onderliggende mechanismen in de plant als reactie op stresssituaties. Op lange termijn zal dit onderzoek ook kunnen bijdragen aan de ontwikkeling van meer stress-resistente planten.

Supplementary data

Supplementary tables

Supplementary table 1 Identification numbers, databases, Pfam ID numbers and Pfam names of the model sequences for each lectin family.

Lectin family	ID model sequence	Database	Pfam ID domain	Pfam domain name
ABA	Q00022.3	UniProtKB	PF07367	FB_lectin
Amaranthin	AAL05954.1	Genbank	PF07468	Agglutinin
CRA	ABL98074.1	Genbank	-	-
Cyanovirin	P81180.2	UniProtKB	PF08881	CVNH
EUL	ABW73993.1	Genbank	-	-
GNA	P30617	UniProtKB	PF01453	B_lectin
Hevein	ABW34946.1	Genbank	PF00187	Chitin_bind_1
Jacalin	AAA32680.1	Genbank	PF01419	Jacalin
Legume lectin	P05046	UniProtKB	PF00139	Lectin_legB
LysM	BAN83772.1	Genbank	PF01476	LysM
Nictaba	AAK84134.1	Genbank	PF14299	PP2
Ricin B	PDB: 2AAI_B	Genbank	PF00652	Ricin_B_lectin

Supplementary table 2 Overview of all putative lectins from *Arabidopsis thaliana* and their protein domains.

AGI code	Protein domains
CRA homologs	
AT4G19720.1	CRA
AT4G19730.1	CRA/CID
AT4G19740.1	CRA
AT4G19750.1	CRA/CID
AT4G19760.1	CRA/CID
AT4G19770.1	CRA/CID
AT4G19800.1	CRA/CID
AT4G19810.1	CRA/CID
AT4G19820.1	CRA/CID
EUL homologs	
AT2G39050.1	EUL
GNA homologs	
AT1G11280.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G11300.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G11330.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G11340.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G11350.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G11410.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK

Supplementary data

AT1G16905.1	GNA
AT1G34300.1	GNA/S-locus glycoprotein/Protein kinase
AT1G61360.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G61370.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61380.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61390.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61400.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61420.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61430.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61440.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61460.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61480.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61490.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61500.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61550.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G61610.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT1G65790.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G65800.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT1G67520.1	GNA/PAN/Protein kinase
AT1G78820.1	GNA/PAN
AT1G78830.1	GNA/PAN
AT1G78850.1	GNA
AT1G78860.1	GNA
AT2G01780.1	GNA
AT2G19130.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT2G41890.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT3G12000.1	GNA/S-locus glycoprotein/PAN
AT3G16030.1	GNA/PAN/Protein kinase
AT3G51710.1	GNA/PAN
AT4G00340.1	GNA/S-locus glycoprotein/Protein kinase
AT4G03230.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT4G11900.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT4G21380.1	GNA/S-locus glycoprotein/PAN/Protein kinase/SRK
AT4G21390.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT4G27290.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT4G27300.1	GNA/S-locus glycoprotein/PAN/Protein kinase
AT4G32300.1	GNA/Protein kinase
AT5G03700.1	GNA/S-locus glycoprotein/PAN
AT5G18470.1	GNA
AT5G24080.1	GNA/S-locus glycoprotein/Protein kinase
AT5G35370.1	GNA/Protein kinase
AT5G39370.1	GNA
AT5G60900.1	GNA/Protein kinase

Hevein homologs	
AT1G56680.1	Hevein/GH19
AT2G43570.1	Hevein/GH19
AT2G43580.1	Hevein/GH19
AT2G43590.1	Hevein/GH19
AT2G43600.1	Hevein/GH19
AT2G43610.1	Hevein/GH19
AT2G43620.1	Hevein/GH19
AT3G04720.1	Hevein/Barwin
AT3G12500.1	Hevein/GH19
AT3G54420.1	Hevein/GH19
JRL homologs	
AT1G05760.1	Jacalin
AT1G05770.1	Jacalin
AT1G19715.1	Jacalin
AT1G33790.1	Jacalin
AT1G52000.1	Jacalin
AT1G52030.1	Jacalin
AT1G52040.1	Jacalin
AT1G52050.1	Jacalin
AT1G52060.1	Jacalin
AT1G52070.1	Jacalin
AT1G52100.1	Jacalin
AT1G52110.1	Jacalin
AT1G52120.1	Jacalin
AT1G52130.1	Jacalin
AT1G57570.1	Jacalin
AT1G58160.1	Jacalin
AT1G60095.1	Jacalin
AT1G60110.1	Jacalin
AT1G60130.1	Jacalin
AT1G61230.1	Jacalin
AT1G73040.1	Jacalin
AT2G25980.1	Jacalin
AT2G33070.1	Jacalin/Kelch1
AT2G39310.1	Jacalin
AT2G39330.1	Jacalin
AT2G43730.1	Jacalin
AT2G43740.1	Jacalin
AT2G43745.1	Jacalin
AT3G16390.1	Jacalin/Kelch1
AT3G16400.1	Jacalin/Kelch1
AT3G16410.1	Jacalin/Kelch1
AT3G16420.1	Jacalin

Supplementary data

AT3G16430.1	Jacalin
AT3G16440.1	Jacalin
AT3G16450.1	Jacalin
AT3G16460.1	Jacalin
AT3G16470.1	Jacalin
AT3G21380.1	Jacalin
AT3G59590.1	Jacalin/F-box associated domain type 1
AT3G59610.1	Jacalin/F-box associated domain type 1/F-box
AT3G59620.1	Jacalin
AT5G28520.1	Jacalin
AT5G35940.1	Jacalin
AT5G35950.1	Jacalin
AT5G38540.1	Jacalin
AT5G38550.1	Jacalin
AT5G46000.1	Jacalin
AT5G49850.1	Jacalin
AT5G49860.1	Jacalin
AT5G49870.1	Jacalin
<u>Legume lectin homologs</u>	
AT5G42120.1	Legume/Protein kinase
AT2G29220.1	Legume/Protein kinase
AT2G29250.1	Legume/Protein kinase
AT2G43700.1	Legume
AT3G15356.1	Legume
AT4G29050.1	Legume/Protein kinase
AT4G02410.1	Legume/Protein kinase
AT3G09190.1	Legume
AT5G01090.1	Legume
AT5G60270.1	Legume/Protein kinase
AT3G45420.1	Legume/Protein kinase
AT5G01540.1	Legume/Protein kinase
AT5G10530.1	Legume/Protein kinase
AT4G02420.1	Legume/Protein kinase
AT3G59700.1	Legume/Protein kinase
AT3G45330.1	Legume/Protein kinase
AT3G53380.1	Legume/Protein kinase
AT5G59270.1	Legume/Protein kinase
AT1G53080.1	Legume
AT1G53070.1	Legume
AT1G15530.1	Legume/Protein kinase
AT4G28350.1	Legume/Protein kinase
AT1G53060.1	Legume
AT5G03140.1	Legume/Protein kinase
AT5G06740.1	Legume/Protein kinase

AT3G08870.1	Legume/Protein kinase
AT4G04960.1	Legume/Protein kinase
AT3G09035.1	Legume
AT3G55550.1	Legume/Protein kinase
AT5G60280.1	Legume/Protein kinase
AT5G60300.1	Legume/Protein kinase
AT3G45440.1	Legume/Protein kinase
AT2G37710.1	Legume/Protein kinase
AT3G53810.1	Legume/Protein kinase
AT3G59750.1	Legume/Protein kinase
AT5G01550.1	Legume/Protein kinase
AT3G59740.1	Legume/Protein kinase
AT2G43690.1	Legume
AT5G55830.1	Legume/Protein kinase
AT3G54080.1	Legume
AT1G70130.1	Legume/Protein kinase
AT3G16530.1	Legume
AT1G70110.1	Legume/Protein kinase
AT5G03350.1	Legume
AT3G59730.1	Legume/Protein kinase
AT3G45390.1	Legume/Protein kinase
AT3G45410.1	Legume/Protein kinase
AT5G60310.1	Legume/Protein kinase
AT5G60320.1	Legume/Protein kinase
AT5G59260.1	Legume/Protein kinase
AT3G45430.1	Legume/Protein kinase
AT1G07460.1	Legume
AT5G65600.1	Legume/Protein kinase
AT5G01560.1	Legume/Protein kinase
<u>LysM homologs</u>	
AT1G21880.1	LysM
AT1G51940.1	LysM/Protein kinase
AT1G55000.1	LysM/F-box
AT1G77630.1	LysM
AT2G17120.1	LysM
AT2G23770.1	LysM/Protein kinase
AT2G33580.1	LysM/Protein kinase
AT3G01840.1	LysM/Protein kinase
AT3G21630.1	LysM/Protein kinase
AT3G52790.1	LysM
AT4G25433.1	LysM
AT5G23130.1	LysM
<u>Nictaba homologs</u>	
AT1G09155.1	Nictaba/F-box

Supplementary data

AT1G10155.1	Nictaba
AT1G12710.1	Nictaba/F-box
AT1G31200.1	Nictaba
AT1G33920.1	Nictaba
AT1G56240.1	Nictaba/F-box
AT1G56250.1	Nictaba/F-box
AT1G63090.1	Nictaba/F-box
AT1G65390.1	Nictaba/TIR
AT1G80110.1	Nictaba/F-box
AT2G02230.1	Nictaba/F-box
AT2G02240.1	Nictaba/F-box
AT2G02250.1	Nictaba/F-box
AT2G02280.1	Nictaba
AT2G02300.1	Nictaba/F-box
AT2G02310.1	Nictaba/F-box
AT2G02320.1	Nictaba/F-box
AT2G02340.1	Nictaba/F-box
AT2G02350.1	Nictaba/F-box
AT2G02360.1	Nictaba/F-box
AT2G26820.1	Nictaba/AIG1-type G
AT3G53000.1	Nictaba/F-box
AT3G61060.2	Nictaba/F-box
AT4G19840.1	Nictaba
AT4G19850.1	Nictaba
AT5G24560.1	Nictaba/F-box
AT5G45070.1	Nictaba/TIR
AT5G45080.1	Nictaba/TIR
AT5G45090.1	Nictaba/TIR
AT5G52120.1	Nictaba/F-box
Ricin B homologs	
AT1G13130.1	Ricin B/GH5
AT3G26140.1	Ricin B/GH5

Supplementary table 3 Literature related to putative lectins from *Arabidopsis thaliana*.

Lectin name	AGI code	Function	Reference
CRA homologs			
AtChiC	AT4G19810	Upregulated expression after ABA, JA and salt treatment.	Ohnuma <i>et al.</i> , 2011
EUL homologs			
ArathEULS3	AT2G39050	Upregulated expression after glutathione treatment.	Hacham <i>et al.</i> , 2014
		Upregulated expression after MeJA, ABA and salt treatment.	Van Hove <i>et al.</i> , 2014
		Role in ABA-induced stomatal closure, upregulated expression after <i>P. syringae</i> infection and overexpression of ArathEULS3 makes plants more tolerant to <i>P. syringae</i> .	Van Hove <i>et al.</i> , 2015
GNA homologs			
CBRLK1	AT1G11350	Ca ²⁺ -dependent CaM binding.	Kim <i>et al.</i> , 2009
-	AT1G61360	Co expression with BAK1-interacting RLK2.	Blaum <i>et al.</i> , 2014
-	AT1G65790	Upregulated expression after infection with <i>F. oxysporum</i> .	Zhu <i>et al.</i> , 2013
-	AT4G21390	Upregulated expression after ozone treatment.	Xu <i>et al.</i> , 2015
-	AT5G18470	Upregulated in plants exposed to lipopolysaccharides.	Sanabria <i>et al.</i> , 2008
-	AT5G60900	Upregulated expression after infection with <i>F. oxysporum</i> .	Zhu <i>et al.</i> , 2013
		Upregulated in plants exposed to lipopolysaccharides.	Sanabria <i>et al.</i> , 2008
Hevein homologs			
PR-4	AT3G04720	Upregulated expression after MeJA and ET treatment.	Thomma <i>et al.</i> , 1999
		Significantly induced by high salt stress and plays as such a role in germination when there is salt stress.	Seo <i>et al.</i> , 2008
		Involved in defence mechanism against <i>A. brassicicola</i> .	Mukherjee <i>et al.</i> , 2010
		Upregulated expression after sulfur dioxide exposure.	Li and Yi, 2012
PR-3	AT3G12500	Upregulated expression after MeJA and ET treatment.	Thomma <i>et al.</i> , 1999
		Significantly induced by high salt stress.	Seo <i>et al.</i> , 2008
		Involved in defence mechanism against <i>A. brassicicola</i> .	Mukherjee <i>et al.</i> , 2010
		Upregulated expression after sulfur dioxide exposure.	Li and Yi, 2012

Supplementary data

AtEP3 AtchitIV	AT3G54420	Induced expression after NO treatment.	Parani <i>et al.</i> , 2004
		Plays a role in programmed cell death.	Passarinho <i>et al.</i> , 2001
JRL homologs			
RTM1	AT1G05760	Restricts the long-distance movement of TEV in the phloem.	Chisholm <i>et al.</i> , 2001
-	AT1G52000	Upregulated expression after infection with <i>F. oxysporum</i> .	Zhu <i>et al.</i> , 2013
MBP2	AT1G52030	Myrosinase binding proteins.	Capella <i>et al.</i> , 2001
MBP1	AT1G52040	Myrosinase binding proteins.	Capella <i>et al.</i> , 2001
AtNSP2	AT2G33070	Nitril specifier proteins.	Kong <i>et al.</i> , 2012
AtNSP3	AT3G16390	Nitril specifier proteins.	Kong <i>et al.</i> , 2012
AtNSP1	AT3G16400	Nitril specifier proteins.	Kong <i>et al.</i> , 2012
AtNSP4	AT3G16410	Nitril specifier proteins.	Kong <i>et al.</i> , 2012
PBP1	AT3G16420	Induced expression after infection with <i>O. ramosa</i> (jasmonate dependent).	Dos Santos <i>et al.</i> , 2003
		Involved in development of chloroplasts.	de Luna-Valdez <i>et al.</i> , 2014
		Plays a role in response to tissue damage.	Nagano <i>et al.</i> , 2005
AtJAC1	AT3G16470	Plays a role in controlling the flowering time.	Xiao <i>et al.</i> , 2015
Legume lectin homologs			
LecRK-b2	AT1G70130	Induced expression after ABA, salt and osmotic stress treatment.	Deng <i>et al.</i> , 2009
-	AT3G15350	Upregulated expression after ozone treatment.	Xu <i>et al.</i> , 2015
		Induced expression after infection with <i>A. brassicicola</i> .	Mukherjee <i>et al.</i> , 2010
-	AT3G16530	Upregulated expression after ozone treatment.	Xu <i>et al.</i> , 2015
AtLecRK2	AT3G45410	Upregulated expression after salt treatment (ET dependent).	He <i>et al.</i> , 2004
SGC Lectin RLK	AT3G53810	Plays a role in pollen development.	Wan <i>et al.</i> , 2008
-	AT3G59700	Upregulated expression after infection with <i>F. oxysporum</i> .	Zhu <i>et al.</i> , 2013
AtLPK1	AT4G02410	Highly upregulated expression after ABA, MeJA and SA treatments.	Huang <i>et al.</i> , 2013
		Contributes to resistance against <i>B. cinerea</i> .	Huang <i>et al.</i> , 2013
LecRK-VI.2/ LecRKA4.1	AT5G01540	Plays a role in the ABA stress response during seed germination.	Xin <i>et al.</i> , 2009
		Upregulated expression after infection with <i>F. oxysporum</i> .	Zhu <i>et al.</i> , 2013
		Contributes to resistance against <i>P. syringae</i> en <i>P. carotovorum</i> .	Singh <i>et al.</i> , 2012
LecRKA4.2	AT5G01550	Plays a role in the ABA stress response during seed germination.	Xin <i>et al.</i> , 2009
LecRKA4.3	AT5G01560	Plays a role in the ABA stress response during seed germination.	Xin <i>et al.</i> , 2009

LecRK-I.9	AT5G60300	Contributes to resistance against <i>P. brassicae</i> .	Bouwmeester <i>et al.</i> , 2011
		Binds proteins that contain RGD motifs.	Gouget <i>et al.</i> , 2006
LysM homologs			
AtCERK1/ AtLYK1/ LysM RLK1	AT3G21630	Important in the response of Arabidopsis to fungi.	Wan <i>et al.</i> , 2008
AtLYK3	AT1G51940	Cross talk between ABA and pathogen stress response.	Paparella <i>et al.</i> , 2014
AtLYK4	AT2G23770	Involved in chitin signal transduction and plant innate immunity.	Wan <i>et al.</i> , 2012
AtLYK5	AT2G33580	Potential involvement in chitin signaling, chitin binding.	Wan <i>et al.</i> , 2012
Nictaba homologs			
VBF	AT1G56250	Makes Arabidopsis more tolerant to <i>Agrobacterium tumefaciens</i> infection.	Zaltsman <i>et al.</i> , 2010
-	AT3G61060	Downregulated expression during callus initiation.	Xu <i>et al.</i> , 2012
PP2-A1	AT4G19840	Binds to <i>N</i> -acetylglucosamine oligomers, high-mannose <i>N</i> -glycans and 9-acyl- <i>N</i> -acetylneuraminic sialic acid.	Beneteau <i>et al.</i> , 2010
		Expression induced after pathogen attack and ET treatment.	Lee <i>et al.</i> , 2014
		Shows molecular chaperone as well as antifungal activity.	Lee <i>et al.</i> , 2014
PP2-B11	AT1G80110	Overexpression lines are more tolerant to high salinity conditions.	Jia <i>et al.</i> , 2015
		Overexpression lines are more sensitive to drought stress.	Lee <i>et al.</i> , 2014
PP2-B10/ F-box Nictaba	AT2G02360	Specific for <i>N</i> -acetylglucosamine, Lewis A, Lewis X, Lewis Y and blood type B motifs.	Stefanowicz <i>et al.</i> , 2012
		Overexpression lines are more tolerant to infection with <i>P. syringae</i> .	Stefanowicz <i>et al.</i> , 2016
Ricine B homologs			
-	AT3G26140	Upregulated expression after infection with the PPV.	Babu <i>et al.</i> , 2008

Supplementary table 4 Gene specific primers for cloning.

Target gene	Forward primer (5'-3')	Reverse primer (5'-3')
AN3 (AT4G19850.2)	evd721 ATGGGAATAATATGGTCTATC	evd722 TCATGCCTCGTGACATAAATC
AN4 (AT1G31200)	evd723 ATGTCTTCACAAAAGAGTTCGC	evd724 TTACACTTCTTGAACAAAGGC
AN5 (AT4G19840)	evd725 ATGAGCAAGAAACATTGCTCAG	evd726 TTACTGTTTGGGACGAATTGC

Supplementary table 5 qPCR primers.

Target gene	Info	Forward primer (5'-3')	Reverse primer (5'-3')
<i>AN3</i> (AT4G19850.2)	/	evd844 TCTTCTCAAAGACAAAGCCACA	evd845 GCTTCAAGGAAAAGTCATCGTC
<i>AN4</i> (AT1G31200)	/	evd846 TTCACAAAAGAGTTCGCATCA	evd847 GAGTCACCTCCCCAAACAAA
<i>AN5</i> (AT4G19840)	/	evd848 GCCACCGGTGACAACCTTAC	evd849 GAGGAGAGAGAGATCGGAGGA
<i>PP2A</i> (AT1G13320)	Reference gene for data normalization	evd727 TCCGAGATCACATGTTCCAAACTC	evd728 CCGTATCATGTTCTCCACAACCG
<i>TIP41</i> (AT4G34270)	Reference gene for data normalization	evd729 TGAAGTGGCTGACAATGGAGTG	evd730 CATGAGCTTGGCATGACTCTCAC
<i>UBC9</i> (AT4G27960)	Reference gene for data normalization	evd731 TCCTACTTCATGTAGCGCAGGAC	evd732 TCCTCCAGAATAAGGGCTATCCG
<i>JMT</i> (AT1G19640)	Positive control for MeJA treatment	evd745 TATGTAAGCTCGCCACGATACGCT	evd746 AACACGATCAACCGGCTCTAACGA
<i>Cor15A</i> (AT2G42540)	Positive control for ABA treatment	evd781 CAGTCAAACCGCAGATACATTGG	evd782 GGCTTCTTTTCCTTTCTCCTCC
<i>WRKY70</i> (AT3G56400)	Positive control for SA treatment	evd811 CATGGATTCCGAAGATCACA	evd812 CTGGCCACACCAATGACAA
<i>RD29A</i> (AT5G52310)	Positive control for NaCl treatment	evd749 ATCACTTGGCTCCACTGTTGTTTC	evd750 ACAAAACACACATAAACATCCAAA GT
<i>Hsp70b</i> (AT1G16030)	Positive control for heat treatment	evd735 ATGTATCAGGGTGGTGCTGCT	evd736 ACCTCTTCGATCTTGGGACCT
<i>PR1</i> (AT2G14610)	Positive control for <i>P. syringae</i> and <i>M. persicae</i> stress	evd1019 GCTACGCAGAACAACCTAAGAGG	evd1020 GCCTTCTCGCTAACCCACAT
<i>PDF1.2</i> (AT5G44420)	Positive control for <i>B. cinerea</i> infection	evd788 AAGTTGTGCGAGAAGCCAAG	evd789 CCATGTTTGGCTCCTTCAAG

<i>ACT2</i> (AT3G18780)	Reference gene for <i>P. syringae</i> biomass	P112 GATGAGGCAGGTCCAGGAATC	P113 GTTTGTACACACAAGTGCATC
<i>PEX4</i> (AT5G25760)	Reference gene for <i>P. syringae</i> biomass	P116 TGCAACCTCCTCAAGTTCG	P117 CACAGACTGAAGCGTCCAAG
<i>oprf</i> (PSPTO_2299)	<i>Pseudomonas</i> gene for <i>P. syringae</i> biomass	P508 AACTGAAAAACACCTTGGGC	P509 CCTGGGTTGTTGAAGTGTA

Supplementary table 6 EGFP fusion construct primers.

Target gene	Construct	Forward primer (5'-3') ^a	Reverse primer (5'-3') ^b
<i>AN3</i> (AT4G19850.2)	EGFP-AN3	evd850 <u>AAAAAGCAGGCTTC</u> ATGGGAATAA TATGGTCTATCTTC	evd851 AGAAAGCTGGGTGTCATGCCTCGTGT ACATAAATC
<i>AN3</i> (AT4G19850.2)	AN3-EGFP	evd852 <u>AAAAAGCAGGCTTC</u> ACCATGGGAA TAATATGGTCTATC	evd853 AGAAAGCTGGGTGTCCTCGTGTACA TAAATC
<i>AN4</i> (AT1G31200)	EGFP-AN4	evd769 <u>AAAAAGCAGGCTTC</u> ATGTCTTCAC AAAAGAGTTCGC	evd770 AGAAAGCTGGGTGTTACACTTCTTGA ACAAAGGCTTC
<i>AN4</i> (AT1G31200)	AN4-EGFP	evd771 <u>AAAAAGCAGGCTTC</u> ACCATGTCTT CACAAAAGAGTTC	evd772 AGAAAGCTGGGTGCACTTCTTGAACA AAGGCTTCG
<i>AN5</i> (AT4G19840)	EGFP-AN5	evd773 <u>AAAAAGCAGGCTTC</u> ATGAGCAAG AAACATTGCTCAG	evd774 AGAAAGCTGGGTGTTACTGTTTGGGA CGAATTGC
<i>AN5</i> (AT4G19840)	AN5-EGFP	evd775 <u>AAAAAGCAGGCTTC</u> ACCATGAGCA AGAAACATTGCTC	evd776 AGAAAGCTGGGTGCTGTTTGGGACGA ATTGC
attB1- <i>ArathNictaba</i> -attB2		evd2 GGGGACAAGTTTGTACAAAAAAG CAGGCT	evd4 GGGGACAAGTTTGTACAAAAAAGCAG GCT

^a Underlined nucleotides are the first part of the attB1 gateway cloning site.

^b Underlined nucleotides are the reverse complementary of the first part of the attB2 gateway cloning site.

Supplementary table 7 Primers for transgenic T-DNA insertion and overexpression lines.

Target (gene)	Info	Forward primer (5'-3')	Reverse primer (5'-3')
<i>ACT2</i> (AT3G18780)	Check for gDNA quality	evd280 GGCTGGATTTGCTGGAGATGAT GC	evd281 GTACGACCACTGGCATACAGGGA
<i>NptII</i> (pO86A1_p160)	Kanamycin resistance gene	evd463 GAACAAGATGGATTGCACGCAG G	evd261 TCAGAAGAAGCTCGTCAAGAAGGCG
T-DNA insertion (SALK line)	Left border primer on T- DNA insertion of SALK line	/	P99 CCTGGGTTGTTGAAGTGGTA
T-DNA insertion (SAIL line)	Left border primer on T- DNA insertion of SAIL line	/	P538 TAGCATCTGAATTCATAACCAATC TCGATACAC
T-DNA insertion line <i>AN4</i> (SALK_019483)	Primers spanning the T-DNA insertion site	P537 (RP) GCTCAAACAGAGCAACGATC	P536 (LP) ACCGATTCGATATGTTTTCCC
T-DNA insertion line <i>AN5</i> (SAIL_835_C05)	Primers spanning the T-DNA insertion site	P540 (RP) TGGGATTACGATAGTTGCTGC	P539 (LP) GGAAGTTCCTCAAACCTCACC
<i>ArathNictaba</i> overexpression constructs	35S promoter until 35S terminator	evd472 GAAACCTCCTCGGATTCCAT	P1 AGGTCACCTGGATTTTGGTTT

Supplementary table 8 Protein expression construct primers.

Target	Construct	Forward primer (5'-3') ^a	Reverse primer (5'-3') ^a
AN3 (AT4G19850.2)	Gibson- His6-Gly3- AN3- Gibson	P266 ATGCACCATCACCATCACCATGGT GGTGGTGGGAATAATATGGTCTATC TTCTCAA	P267 <u>GCTTTGTTAGCAGCCGGATCTCATGCC</u> TCGTGTACATAAATC
		P213 <u>TTAAGAAGGAGATATACGGGATG</u> CACCATCACCATCACCATG	P267
AN3-Gly3-His6 (AT4G19850.2)	Gibson- AN3-Gly3- His6- Gibson	P268 <u>TTAAGAAGGAGATATACGGGATG</u> GGAATAATATGGTCTATCTTCTC	P216 <u>GCTTTGTTAGCAGCCGGATCTCAATG</u> ATGATGATGATGATGTCC
AN4 (AT1G31200)	Gibson- His6-Gly3- AN4- Gibson	P239 ATGCACCATCACCATCACCATGGT GGTGGTCTTTCACAAAAGAGTTTCG CATC	P240 <u>GCTTTGTTAGCAGCCGGATCTTACTACT</u> TCTTGAACAAAAGGCTTC
		P213 <u>TTAAGAAGGAGATATACGGGATG</u> CACCATCACCATCACCATG	P240
AN4-Gly3-His6 (AT1G31200)	Gibson- AN4-Gly3- His6- Gibson	P241 <u>TTAAGAAGGAGATATACGGGATG</u> TCTTACAAAAGAGTTTCGC	P216 <u>GCTTTGTTAGCAGCCGGATCTCAATG</u> ATGATGATGATGATGTCC
AN5 (AT4G19840)	Gibson- His6-Gly3- AN5- Gibson	P263 ATGCACCATCACCATCACCATGGT GGTGGTAGCAAGAAACATTGCTCA GAATTAT	P264 <u>GCTTTGTTAGCAGCCGGATCTTACTGT</u> TTGGGACGAATTGC
		P213 <u>TTAAGAAGGAGATATACGGGATG</u> CACCATCACCATCACCATG	P264
AN5-Gly3-His6 (AT4G19840)	Gibson- AN5-Gly3- His6- Gibson	P265 <u>TTAAGAAGGAGATATACGGGATG</u> AGCAAGAAACATTGCTC	P216 <u>GCTTTGTTAGCAGCCGGATCTCAATG</u> ATGATGATGATGATGTCC
pET21a(+) vector		P242 GATCCGGCTGCTAACAAAG	P243 CCCGTATATCTCCTTCTTAAAG
pET21a(+) vector (sequencing)		P258 TAATACGACTCACTATAGGG	P259 AAAGGGAATAAGGGCGACAC

^a Underlined nucleotides are the Gibson assembly sites.

Supplementary table 9 Significant ($p \leq 0.01$) MS hits comparing the pull-down assays using induced (three replicates) and non-induced (three replicates) soluble *E. coli* fractions (I/NI), and the plant lysate from non-treated *Arabidopsis* plants (Figure 5.7A).

Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
hisB	Histidine biosynthesis bifunctional protein HisB	A0A140NAY3	<i>E. coli</i>	2,75	5,84
ibpA	Small Hsp IbpA	A0A140N1Q5	<i>E. coli</i>	2,40	5,54
rpsK	30S ribosomal protein S11	A0A140N7L9	<i>E. coli</i>	1,76	5,50
ECBD_0490	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	A0A140N752	<i>E. coli</i>	2,25	4,86
ibpB	Small Hsp IbpB	A0A140N3G6	<i>E. coli</i>	3,15	4,75
ileS	Isoleucine--tRNA ligase	A0A140ND98	<i>E. coli</i>	2,26	4,68
ECBD_2906	DNA-directed RNA polymerase	A0A140NCE7	<i>E. coli</i>	3,73	4,42
lacZ	Beta-galactosidase	A0A140NDI2	<i>E. coli</i>	2,95	3,86
valS	Valine--tRNA ligase	A0A140NGV6	<i>E. coli</i>	3,04	3,24
topA	DNA topoisomerase 1	A0A140NCX5	<i>E. coli</i>	2,70	2,87
ECBD_1336	Beta-ketoacyl synthase	A0A140N9G9	<i>E. coli</i>	3,39	2,77
ECBD_1047	Uncharacterized protein	A0A140N7A6	<i>E. coli</i>	1,93	-3,85
ECBD_3882	Entericidin EcnAB	A0A140NGM2	<i>E. coli</i>	2,51	-7,45

Supplementary table 10 Significant ($p \leq 0.01$) MS hits comparing the pull-down assays using induced (three replicates) and non-induced (three replicates) soluble *E. coli* fractions (I/NI), and the plant lysate from NaCl-treated Arabidopsis plants (Figure 5.7B).

Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
lbpA	Small Hsp	A0A140N1Q5	<i>E. coli</i>	3,63	6,78
rpsK	30S ribosomal protein S11	A0A140N7L9	<i>E. coli</i>	3,10	5,86
ibpB	Small Hsp IbpB	A0A140N3G6	<i>E. coli</i>	3,42	5,15
ECBD_2906	DNA-directed RNA polymerase	A0A140NCE7	<i>E. coli</i>	4,62	4,69
ECBD_0490	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	A0A140N752	<i>E. coli</i>	3,24	4,56
hisB	Histidine biosynthesis bifunctional protein HisB	A0A140NAY3	<i>E. coli</i>	3,44	4,39
hisD	Histidinol dehydrogenase	A0A140N5W6	<i>E. coli</i>	3,72	3,95
lacZ	Beta-galactosidase	A0A140NDI2	<i>E. coli</i>	3,48	3,91
fabZ	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	A0A140NFC7	<i>E. coli</i>	1,94	3,76
hsI0	33 kDa chaperonin	A0A140N2J1	<i>E. coli</i>	2,11	3,67
hisC	Histidinol-phosphate aminotransferase	A0A140N8D8	<i>E. coli</i>	2,53	3,61
ileS	Isoleucine--tRNA ligase	A0A140ND98	<i>E. coli</i>	1,66	3,60
hsIV	ATP-dependent protease subunit HsIV	A0A140NHQ8	<i>E. coli</i>	3,87	3,39
ECBD_0180	Cold-shock DNA-binding domain protein	A0A140N4F2	<i>E. coli</i>	1,44	3,15
ECBD_3970	Transcriptional regulator, AraC family	A0A140NHD9	<i>E. coli</i>	2,47	3,15
dnaJ	Chaperone protein DnaJ	A0A140NfZ9	<i>E. coli</i>	3,97	2,99
ygiQ	UPF0313 protein YgiQ	A0A140N6D4	<i>E. coli</i>	3,91	2,85
PP2-A9	Protein PP2-LIKE A9	Q9SA16	Recombinant protein	3,81	2,82

Supplementary data

Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
ECBD_1042	Glycine betaine/L-proline ABC transporter, ATPase subunit	A0A140N655	<i>E. coli</i>	1,93	2,82
valS	Valine--tRNA ligase	A0A140NGV6	<i>E. coli</i>	1,91	2,77
BGLU23	Beta-glucosidase 23	Q9SR37	<i>A. thaliana</i>	4,17	2,76
topA	DNA topoisomerase 1	A0A140NCX5	<i>E. coli</i>	2,51	2,70
rfaH	Transcription antitermination protein RfaH	A0A140NI00	<i>E. coli</i>	2,46	2,67
TGG1	Myrosinase 1	P37702	<i>A. thaliana</i>	1,67	2,64
ECBD_1336	Beta-ketoacyl synthase	A0A140N9G9	<i>E. coli</i>	2,14	2,63
yqgE	UPF0301 protein YqgE	A0A140N8C3	<i>E. coli</i>	2,13	2,60
ECBD_4231	Uncharacterized protein	A0A140NDV6	<i>E. coli</i>	2,54	2,60
ECBD_1155	Iron-sulfur cluster assembly scaffold protein IscU	A0A140N923	<i>E. coli</i>	2,97	2,50
tolB	Protein TolB	A0A140NEI9	<i>E. coli</i>	2,31	2,50
hisH	Imidazole glycerol phosphate synthase subunit HisH	A0A140N8L8	<i>E. coli</i>	2,07	2,33
mutM	Formamidopyrimidine-DNA glycosylase	A0A140N6H0	<i>E. coli</i>	2,76	2,22
ybaB	Nucleoid-associated protein YbaB	A0A140NF80	<i>E. coli</i>	3,02	2,20
ECBD_2317	Aldehyde Dehydrogenase	A0A140NAG2	<i>E. coli</i>	4,95	-1,88
ECBD_2279	Amidohydrolase	A0A140N7L1	<i>E. coli</i>	3,56	-1,92
ECBD_1392	Uncharacterized protein	A0A140NAA6	<i>E. coli</i>	2,54	-2,05
ECBD_0225	Transcriptional regulator, AraC family	A0A140N6E1	<i>E. coli</i>	2,66	-2,12
ECBD_2590	NAD(P)H dehydrogenase (quinone)	A0A140N993	<i>E. coli</i>	4,48	-2,15
ECBD_3998	Maltose operon periplasmic	A0A140NFH4	<i>E. coli</i>	3,86	-2,19
nanA	<i>N</i> -acetylneuraminase lyase	A0A140N3N8	<i>E. coli</i>	2,16	-2,21
ECBD_2540	Uncharacterized protein	A0A140N946	<i>E. coli</i>	2,10	-2,31

Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
ECBD_0129	PTS system, mannitol-specific IIC subunit	A0A140N201	<i>E. coli</i>	2,40	-2,34
ECBD_1920	Fructosamine kinase	A0A140N6M2	<i>E. coli</i>	4,35	-2,37
rssB	Regulator of RpoS	A0A140NAZ4	<i>E. coli</i>	4,07	-2,42
aceK	Isocitrate dehydrogenase kinase/phosphatase	A0A140NHI6	<i>E. coli</i>	3,45	-2,48
ECBD_1096	GCN5-related N-acetyltransferase	A0A140N551	<i>E. coli</i>	2,93	-2,52
ECBD_3227	Cytochrome o ubiquinol oxidase, subunit I	A0A140NA88	<i>E. coli</i>	3,38	-2,59
ECBD_0868	PAS modulated sigma54 specific transcriptional regulator, Fis family	A0A140N695	<i>E. coli</i>	1,77	-2,66
ECBD_1962	SufBD protein	A0A140N985	<i>E. coli</i>	1,83	-2,68
ECBD_1047	Uncharacterized protein	A0A140N7A6	<i>E. coli</i>	3,38	-2,84
ECBD_2951	GTP cyclohydrolase 1 type 2 homolog	A0A140NC19	<i>E. coli</i>	2,76	-2,88
ECBD_3372	Molybdopterine dehydrogenase FAD-binding	A0A140NBF3	<i>E. coli</i>	3,19	-3,06
ECBD_3893	Anaerobic C4-dicarboxylate transporter	A0A140SS54	<i>E. coli</i>	1,90	-3,10
acpP	Acyl carrier protein	A0A140NCR5	<i>E. coli</i>	2,50	-3,10
ECBD_1822	PTS system, mannose/fructose/sorbose family, IID subunit	A0A140N9E9	<i>E. coli</i>	2,45	-3,10
ECBD_3373	Aldehyde oxidase and xanthine dehydrogenase molybdopterine binding	A0A140NDP2	<i>E. coli</i>	2,79	-3,12
ECBD_0227	Efflux transporter, RND family, MFP subunit	A0A140N515	<i>E. coli</i>	1,62	-3,19
ECBD_2171	Nitrate reductase, beta subunit	A0A140N9U4	<i>E. coli</i>	2,38	-3,43

Supplementary data

Gene name	Protein name	Protein ID	Origin	-Log p	Log₂(I/NI)
ECBD_2574	Protein PhoH	A0A140NAZ5	<i>E. coli</i>	3,57	-3,46
ECBD_0229	Transcriptional regulator, LuxR family	A0A140N6Z5	<i>E. coli</i>	2,59	-3,53
ECBD_3370	Uncharacterized protein	A0A140NFC3	<i>E. coli</i>	1,64	-3,72
ECBD_2922	Protein TolR	A0A140N9F2	<i>E. coli</i>	1,83	-3,74
ECBD_3148	2-hydroxy-3-oxopropionate reductase	A0A140NC12	<i>E. coli</i>	2,68	-4,74
ECBD_3882	Entericidin EcnAB	A0A140NGM2	<i>E. coli</i>	3,67	-7,00

Supplementary table 11 Significant ($p \leq 0.05$) MS hits comparing the pull-down assays using induced (three replicates) and non-induced (two replicates) soluble *E. coli* fractions (I/NI), and the plant lysate from non-treated *Arabidopsis* plants (Figure 5.7C).

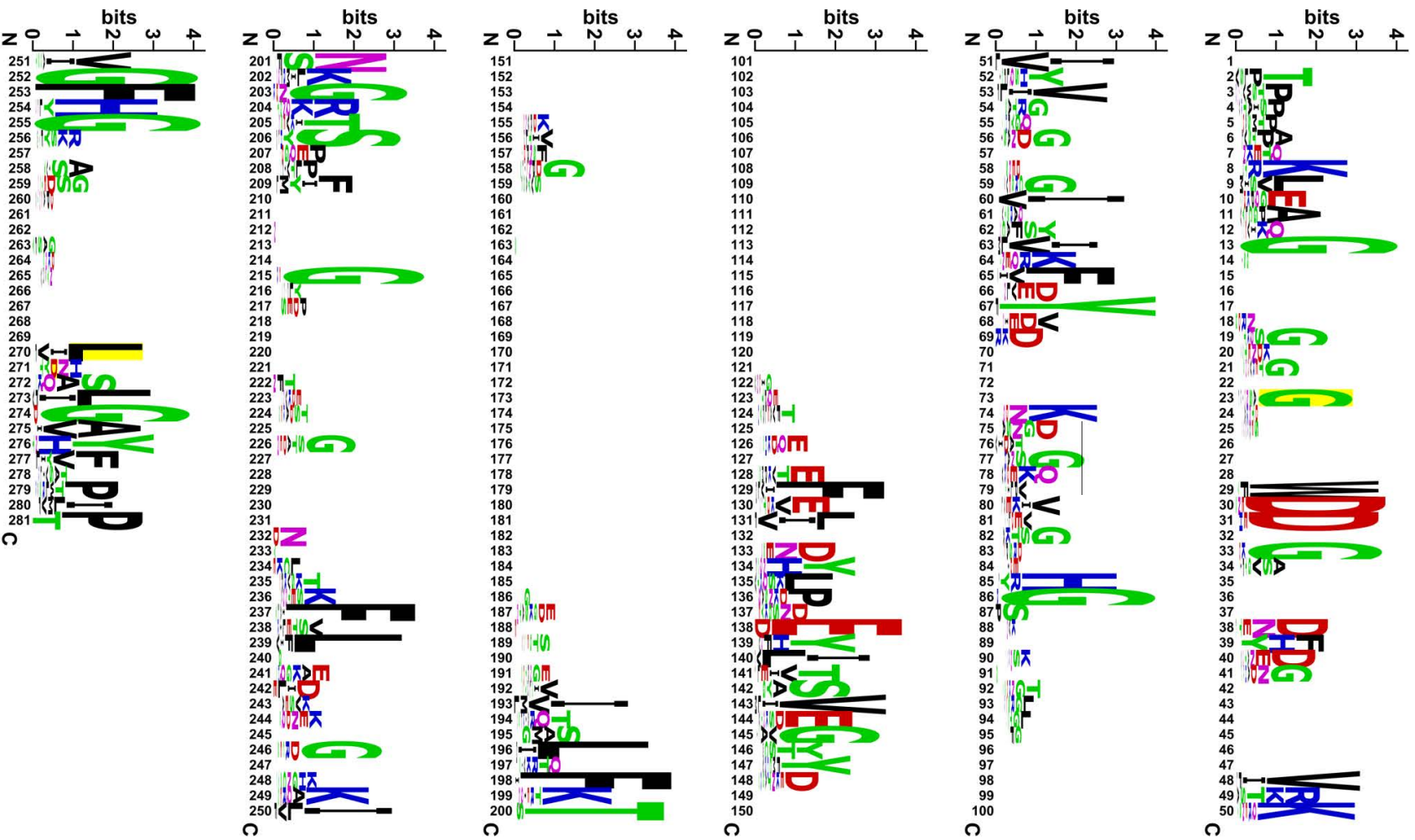
Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
rpsK	30S ribosomal protein S11	A0A140N7L9	<i>E. coli</i>	3,19	6,86
TGG1	Myrosinase 1	P37702	<i>A. thaliana</i>	2,25	6,78
ibpA	Small Hsp IbpA	A0A140N1Q5	<i>E. coli</i>	2,12	6,11
ECBD_0490	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	A0A140N752	<i>E. coli</i>	1,81	5,24
hisB	Histidine biosynthesis bifunctional protein HisB	A0A140NAY3	<i>E. coli</i>	2,64	5,17
hsI0	33 kDa chaperonin	A0A140N2J1	<i>E. coli</i>	1,75	5,16
ESM1	GDSL esterase/lipase ESM1	Q9LJG3	<i>A. thaliana</i>	1,73	5,03
ibpB	Small Hsp IbpB	A0A140N3G6	<i>E. coli</i>	2,27	4,69
ileS	Isoleucine--tRNA ligase	A0A140ND98	<i>E. coli</i>	1,61	4,64
ECBD_2906	DNA-directed RNA polymerase	A0A140NCE7	<i>E. coli</i>	2,73	4,47
hsIV	ATP-dependent protease subunit HsIV	A0A140NHQ8	<i>E. coli</i>	2,40	4,32
lacZ	Beta-galactosidase	A0A140NDI2	<i>E. coli</i>	2,61	4,17
dnaJ	Chaperone protein DnaJ	A0A140NFZ9	<i>E. coli</i>	1,78	3,99
rraB	Regulator of ribonuclease activity B	A0A140NDQ0	<i>E. coli</i>	1,28	3,53
valS	Valine--tRNA ligase	A0A140NGV6	<i>E. coli</i>	2,60	3,47
PP2-A9	Protein PP2-LIKE A9	Q9SA16	Recombinant protein	1,67	3,28
ECBD_0346	Hsp15	A0A140N4B4	<i>E. coli</i>	2,44	3,23
gpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	A0A140N9D9	<i>E. coli</i>	1,98	3,17
yqgE	UPF0301 protein YqgE	A0A140N8C3	<i>E. coli</i>	1,52	3,08
hisD	Histidinol dehydrogenase	A0A140N5W6	<i>E. coli</i>	2,06	3,08
ppa	Inorganic pyrophosphatase	A0A140NEF6	<i>E. coli</i>	1,95	3,04

Supplementary data

Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
ybaB	Nucleoid-associated protein YbaB	A0A140NF80	<i>E. coli</i>	1,43	2,87
topA	DNA topoisomerase 1	A0A140NCX5	<i>E. coli</i>	1,97	2,72
ECBD_1336	Beta-ketoacyl synthase	A0A140N9G9	<i>E. coli</i>	2,46	2,71
lexA	LexA repressor	A0A140NHF7	<i>E. coli</i>	1,51	2,70
PTAC16	Protein plastid transcriptionally active 16, chloroplastic	Q9STF2	<i>A. thaliana</i>	1,74	2,66
hisC	Histidinol-phosphate aminotransferase	A0A140N8D8	<i>E. coli</i>	2,05	2,58
ygiQ	UPF0313 protein YgiQ	A0A140N6D4	<i>E. coli</i>	1,84	2,57
nsrR	HTH-type transcriptional repressor NsrR	A0A140NF33	<i>E. coli</i>	2,03	2,42
mutM	Formamidopyrimidine-DNA glycosylase	A0A140N6H0	<i>E. coli</i>	1,98	2,32
ECBD_2666	Porin Gram-negative type	A0A140NAN5	<i>E. coli</i>	3,95	2,12
ECBD_2991	PhoH family protein	A0A140NAC4	<i>E. coli</i>	2,83	2,00
ECBD_0243	Oligopeptidase A	A0A140N6F6	<i>E. coli</i>	2,83	1,80
ECBD_1563	Putative PTS IIA-like nitrogen-regulatory protein PtsN	A0A140N8Q5	<i>E. coli</i>	2,46	-1,79
ECBD_2279	Amidohydrolase	A0A140N7L1	<i>E. coli</i>	3,13	-1,93
ECBD_1922	Phosphofructokinase	A0A140NB59	<i>E. coli</i>	2,14	-1,98
ECBD_0661	NADH:flavin oxidoreductase/NADH oxidase	A0A140N7Z9	<i>E. coli</i>	2,06	-2,02
ECBD_2590	NAD(P)H dehydrogenase (quinone)	A0A140N993	<i>E. coli</i>	2,73	-2,04
ECBD_2317	Aldehyde Dehydrogenase	A0A140NAG2	<i>E. coli</i>	2,77	-2,08
ECBD_2784	Beta-lactamase	A0A140NE46	<i>E. coli</i>	3,37	-2,22
ECBD_3373	Aldehyde oxidase and xanthine dehydrogenase molybdopterin binding	A0A140NDP2	<i>E. coli</i>	2,02	-2,26

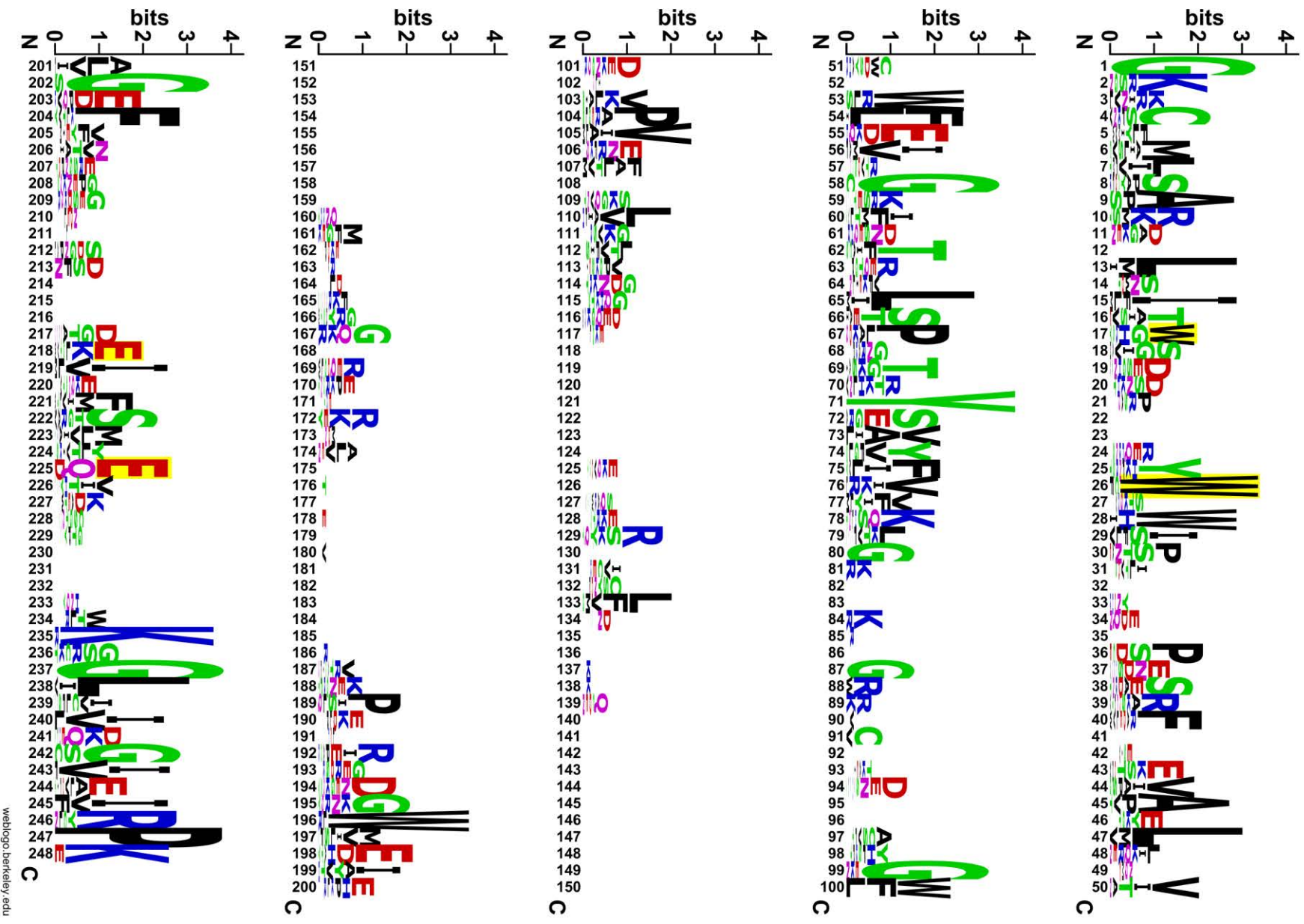
Gene name	Protein name	Protein ID	Origin	-Log p	Log ₂ (I/NI)
rssB	Regulator of RpoS	A0A140NAZ4	<i>E. coli</i>	2,43	-2,26
ECBD_0641	Uncharacterized protein	A0A140N7Y1	<i>E. coli</i>	2,29	-2,27
ECBD_1055	Transcriptional regulator, GntR family	A0A140N8Q4	<i>E. coli</i>	2,33	-2,38
aceK	Isocitrate dehydrogenase kinase/phosphatase	A0A140NHI6	<i>E. coli</i>	2,06	-2,60
ECBD_1744	Trehalose-6-phosphate synthase	A0A140NB73	<i>E. coli</i>	1,58	-2,64
ECBD_0868	PAS modulated sigma54 specific transcriptional regulator, Fis family	A0A140N695	<i>E. coli</i>	1,32	-2,72
ECBD_0229	Transcriptional regulator, LuxR family	A0A140N6Z5	<i>E. coli</i>	1,55	-2,85
ECBD_1096	GCN5-related N-acetyltransferase	A0A140N551	<i>E. coli</i>	1,25	-2,86
ECBD_1054	Peptidoglycan-binding LysM	A0A140N935	<i>E. coli</i>	2,02	-2,90
ECBD_1920	Fructosamine kinase	A0A140N6M2	<i>E. coli</i>	2,43	-2,91
ECBD_2922	Protein TolR	A0A140N9F2	<i>E. coli</i>	1,79	-3,03
ECBD_1532	Transcriptional regulator, MerR family	A0A140N6B8	<i>E. coli</i>	1,50	-3,11
ECBD_3372	Molybdopterin dehydrogenase FAD-binding	A0A140NBF3	<i>E. coli</i>	1,67	-3,11
ECBD_0227	Efflux transporter, RND family, MFP subunit	A0A140N515	<i>E. coli</i>	2,21	-3,13
ECBD_3148	2-hydroxy-3-oxopropionate reductase	A0A140NC12	<i>E. coli</i>	2,18	-3,14
ECBD_1047	Uncharacterized protein	A0A140N7A6	<i>E. coli</i>	1,53	-3,25
ECBD_0225	Transcriptional regulator, AraC family	A0A140N6E1	<i>E. coli</i>	1,82	-3,30
ECBD_2574	Protein PhoH	A0A140NAZ5	<i>E. coli</i>	3,81	-3,49
ECBD_3223	BolA family protein	A0A140NDA3	<i>E. coli</i>	2,16	-3,50
ECBD_3882	Entericidin EcnAB	A0A140NGM2	<i>E. coli</i>	2,24	-6,52

Supplementary figures



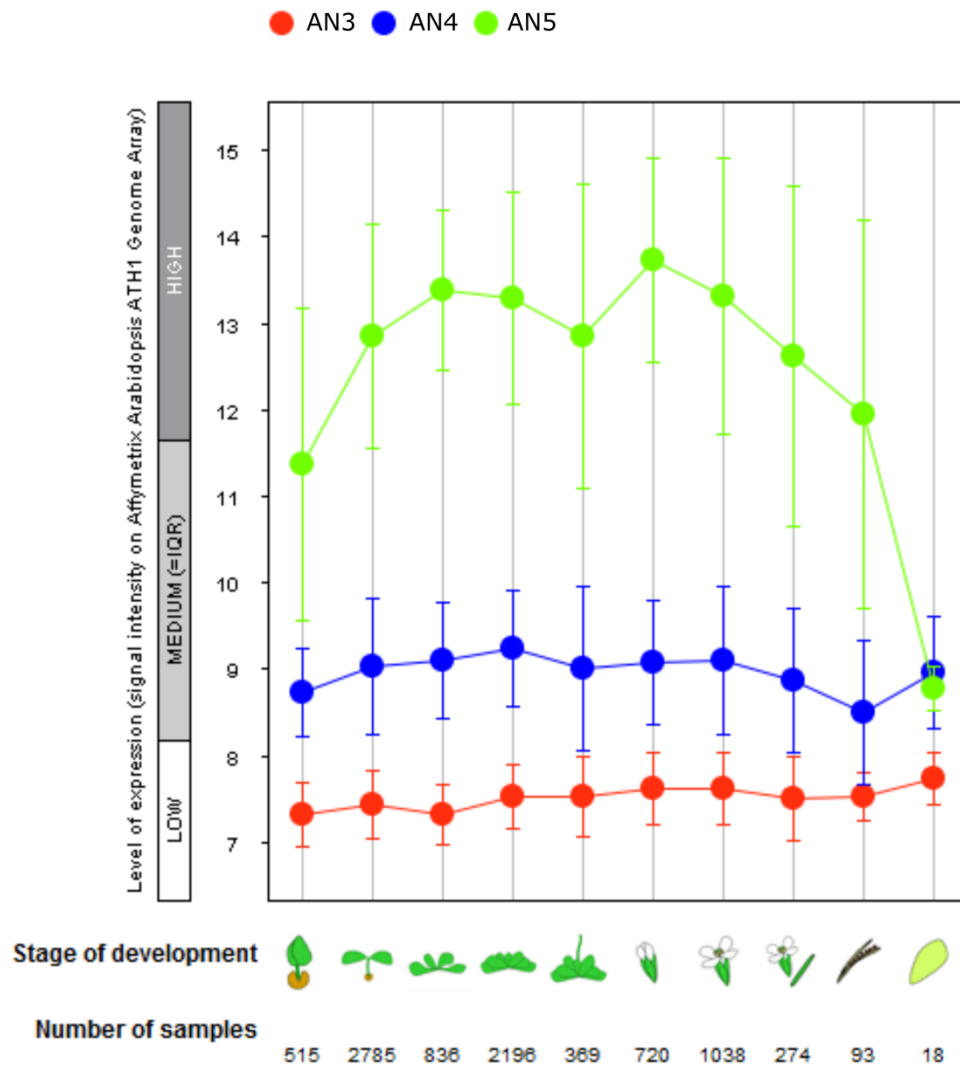
weblogo.berkeley.edu

Supplementary figure 1 Complete Weblogo of the jacalin domains.



Supplementary figure 2 Complete Weblogo of the Nictaba domains.

Dataset: 10 developmental stages from data selection: AT_AFFY_ATH1-1
 Showing 3 measure(s) of 3 gene(s) on selection: *ArathNictabas*



Supplementary figure 3 Expression levels of *ArathNictabas* during development generated with the Genevestigator tool.

References

- Adamová L, Malinovská L, Wimmerová M.** 2014. New sensitive detection method for lectin hemagglutination using microscopy. *Microscopy Research and Technique* **77**, 841–849.
- Adams J, Kelso R, Cooley L.** 2000. The kelch repeat superfamily of proteins: propellers of cell function. *Trends in Cell Biology* **10**, 17–24.
- Ahn YO, Shimizu B, Sakata K, Gantulga D, Zhou C, Zhou Z, Bevan DR, Esen A.** 2010. Scopolin-hydrolyzing β -glucosidases in roots of *Arabidopsis*. *Plant and Cell Physiology* **51**, 132–143.
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P.** 2002. *Molecular biology of the cell*.
- Alexandrov NN, Troukhan ME, Brover V, Tatarinova T, Flavell RB, Feldmann KA.** 2006. Features of *Arabidopsis* genes and genome discovered using full-length cDNAs. *Plant Molecular Biology* **60**, 69–85.
- Alonso JM, Stepanova AN, Leisse TJ, et al.** 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- Andersson D, Chakrabarty R, Bejai S, Zhang J, Rask L, Meijer J.** 2009. Myrosinases from root and leaves of *Arabidopsis thaliana* have different catalytic properties. *Phytochemistry* **70**, 1345–1354.
- Arnold K, Bordoli L, Kopp J, Schwede T.** 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.
- Asamizu E, Nakamura Y, Sato S, Tabata S.** 2000. A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Research* **7**, 175–180.
- Al Atalah B, Fouquaert E, Vanderschaeghe D, Proost P, Balzarini J, Smith DF, Rougé P, Lasanajak Y, Callewaert N, Van Damme EJM.** 2011. Expression analysis of the nucleocytoplasmic lectin ‘Oryzata’ from rice in *Pichia pastoris*. *FEBS Journal* **278**, 2064–2079.
- Al Atalah B, De Vleeschauwer D, Xu J, Fouquaert E, Höfte M, Van Damme EJM.** 2014. Transcriptional behavior of EUL-related rice lectins toward important abiotic and biotic stresses. *Journal of Plant Physiology* **171**, 986–992.
- Atkinson NJ, Urwin PE.** 2012. The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany* **63**, 3523–3544.
- Audenaert K, De Meyer GB, Höfte MM.** 2002. Abscisic acid determines basal susceptibility of tomato to *Botrytis cinerea* and suppresses salicylic acid-dependent signaling mechanisms. *Plant Physiology* **128**, 491–501.
- Babu M, Griffiths JS, Huang T-S, Wang A.** 2008. Altered gene expression changes in *Arabidopsis* leaf tissues and protoplasts in response to plum pox virus infection. *BMC Genomics* **9**, 325.
- Barth C, Jander G.** 2006. *Arabidopsis* myrosinases TGG1 and TGG2 have redundant function in glucosinolate breakdown and insect defense. *The Plant Journal* **46**, 549–562.
- Baxter A, Mittler R, Suzuki N.** 2014. ROS as key players in plant stress signalling. *Journal of Experimental Botany* **65**, 1229–1240.
- Beck EH, Fettig S, Knake C, Hartig K, Bhattarai T.** 2007. Specific and unspecific responses of plants to cold and drought stress. *Journal of Biosciences* **32**, 501–510.
- Beneteau J, Renard D, Marché L, Douville E, Lavenant L, Rahbé Y, Dupont D, Vilaine F, Dinant S.** 2010. Binding properties of the *N*-acetylglucosamine and high-mannose *N*-glycan PP2-A1 phloem lectin in *Arabidopsis*. *Plant Physiology* **153**, 1345–1361.

- Benkert P, Biasini M, Schwede T.** 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350.
- Bhardwaj V, Meier S, Petersen LN, Ingle RA, Roden LC.** 2011. Defence responses of *Arabidopsis thaliana* to infection by *Pseudomonas syringae* are regulated by the circadian clock. *PLoS ONE* **6**, e26968.
- Bhargava S, Sawant K.** 2013. Drought stress adaptation: metabolic adjustment and regulation of gene expression. *Plant Breeding* **132**, 21–32.
- Blaum BS, Mazzotta S, Nöldeke ER, Halter T, Madlung J, Kemmerling B, Stehle T.** 2014. Structure of the pseudokinase domain of BIR2, a regulator of BAK1-mediated immune signaling in *Arabidopsis*. *Journal of Structural Biology* **186**, 112–121.
- Boller T, Felix G.** 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* **60**, 379–406.
- Bortesi L, Fischer R.** 2015. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances* **33**, 41–52.
- Bourne Y, Roig-Zamboni V, Barre A, Peumans WJ, Astoul CH, Van Damme EJM, Rougé P.** 2004. The crystal structure of the *Calystegia sepium* agglutinin reveals a novel quaternary arrangement of lectin subunits with a β -prism fold. *Journal of Biological Chemistry* **279**, 527–533.
- Bouwmeester K, de Sain M, Weide R, Gouget A, Klamer S, Canut H, Govers F.** 2011. The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *PLoS Pathogens* **7**, e1001327.
- Brameier M, Krings A, MacCallum RM.** 2007. NucPred - Predicting nuclear localization of proteins. *Bioinformatics* **23**, 1159–1160.
- Braun P, Aubourg S, Van Leene J, De Jaeger G, Lurin C.** 2013. Plant protein interactomes. *Annual Review of Plant Biology* **64**, 161–187.
- Brouwer M, Lievens B, Van Hemelrijck W, Van Den Ackerveken G, Cammue BPA, Thomma BPHJ.** 2003. Quantification of disease progression of several microbial pathogens on *Arabidopsis thaliana* using real-time fluorescence PCR. *FEMS Microbiology Letters* **228**, 241–248.
- Buist G, Steen A, Kok J, Kuipers OP.** 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Molecular Microbiology* **68**, 838–847.
- Bunn-Moreno MM, Campos-Neto A.** 1981. Lectin(s) extracted from seeds of *Artocarpus integrifolia* (jackfruit): potent and selective stimulator(s) of distinct human T and B cell functions. *Journal of Immunology* **127**, 427–429.
- Burch-Smith TM, Dinesh-Kumar SP.** 2007. The functions of plant TIR domains. *Science Signal Transduction Knowledge Environment* **2007**, pe46.
- Burow M, Zhang ZY, Ober JA, Lambrix VM, Wittstock U, Gershenzon J, Kliebenstein DJ.** 2008. ESP and ESM1 mediate indol-3-acetonitrile production from indol-3-ylmethyl glucosinolate in *Arabidopsis*. *Phytochemistry* **69**, 663–671.
- Büttner D.** 2016. Behind the lines—actions of bacterial type III effector proteins in plant cells. *FEMS Microbiology Reviews* **40**, 894–937.
- Büttner D, He SY.** 2009. Type III protein secretion in plant pathogenic bacteria. *Plant Physiology* **150**, 1656–1664.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T.** 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.

- Capella A, Menossi M, Arruda P, Benedetti C.** 2001. COI1 affects myrosinase activity and controls the expression of two flower-specific myrosinase-binding protein homologues in *Arabidopsis*. *Planta* **213**, 691–699.
- Carrió MM, Villaverde A.** 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology* **96**, 3–12.
- Cayla T, Batailler B, Le Hir R, Revers F, Anstead JA, Thompson GA, Grandjean O, Dinant S.** 2015. Live imaging of companion cells and sieve elements in *Arabidopsis* leaves. *PLoS ONE* **10**, e0118122.
- Chen Y, Peumans W, Hause B, Bras J, Kumar M, Proost P, Barre A, Rougé P, Van Damme EJM.** 2002. Jasmonate methyl ester induces the synthesis of a cytoplasmic/nuclear chitooligosaccharide-binding lectin in tobacco leaves. *FASEB Journal* **16**, 905–907.
- Chisholm ST, Parra MA, Anderberg RJ, Carrington JC.** 2001. *Arabidopsis* RTM1 and RTM2 genes function in phloem to restrict long-distance movement of tobacco etch virus. *Plant Physiology* **127**, 1667–1675.
- Çiftçi YO.** 2012. *Transgenic plants - advances and limitations*.
- Clough SJ, Bent A.** 1998. Floral dip: a simplified method for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**, 735–743.
- Couto D, Zipfel C.** 2016. Regulation of pattern recognition receptor signalling in plants. *Nature Reviews Immunology* **16**, 537–552.
- Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. *Genome Research* **14**, 1188–1190.
- Cui F, Wu S, Sun W, Coaker G, Kunkel B, He P, Shan L.** 2013. The *Pseudomonas syringae* type III effector *AvrRpt2* promotes pathogen virulence via stimulating *Arabidopsis* auxin/indole acetic acid protein turnover. *Plant Physiology* **162**, 1018–1029.
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R.** 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology* **139**, 5–17.
- Dang L.** 2017. Chimeric lectins with toxin domains in cucumber. PhD thesis, Ghent University.
- Dang L, Van Damme EJM.** 2016. Genome-wide identification and domain organization of lectin domains in cucumber. *Plant Physiology and Biochemistry* **108**, 165–176.
- Dangl JL, Horvath DM, Staskawicz BJ.** 2013. Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751.
- Dangl JL, Jones JD.** 2001. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833.
- Dar TA, Uddin M, Khan MMA, Hakeem KR, Jaleel H.** 2015. Jasmonates counter plant stress: a review. *Environmental and Experimental Botany* **115**, 49–57.
- Decroocq V, Sicard O, Alamillo JM, Lansac M, Eyquard JP, García JA, Candresse T, Le Gall O, Revers F.** 2006. Multiple resistance traits control plum pox virus infection in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **19**, 541–549.
- Delporte A.** 2013. Expression analysis and interaction studies of the tobacco lectin. PhD thesis, Ghent University.
- Delporte A, Van Holle S, Lannoo N, Van Damme EJM.** 2015. The tobacco lectin, prototype of the family of Nictaba-related proteins. *Current Protein and Peptide Science* **16**, 5–16.

- Delporte A, Lannoo N, Vandeborre G, Ongenaert M, Van Damme EJM.** 2011. Jasmonate response of the *Nicotiana tabacum* agglutinin promoter in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* **49**, 843–851.
- Delporte A, De Vos WH, Van Damme EJM.** 2014. *In vivo* interaction between the tobacco lectin and the core histone proteins. *Journal of Plant Physiology* **171**, 1149–1156.
- Delseny M, Pelletier G.** 2001. From *Arabidopsis* to rice genomics: a survey of French programmes. *Comptes Rendus de l'Academie des Sciences - Serie III* **324**, 1103–1110.
- de Luna-Valdez LA, Martinez-Batallar AG, Hernandez-Ortiz M, Encarnacion-Guevara S, Ramos-Vega M, Lopez-Bucio JS, Leon P, Guevara-Garcia AA.** 2014. Proteomic analysis of chloroplast biogenesis (clb) mutants uncovers novel proteins potentially involved in the development of *Arabidopsis thaliana* chloroplasts. *Journal of Proteomics* **111**, 148–164.
- Demain AL, Vaishnav P.** 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* **27**, 297–306.
- Deng K, Wang Q, Zeng J, Guo X, Zhao X, Tang D, Liu X.** 2009. A lectin receptor kinase positively regulates ABA response during seed germination and is involved in salt and osmotic stress response. *Journal of Plant Biology* **52**, 493–500.
- Dezfulian MH, Soulliere DM, Dhaliwal RK, Sareen M, Crosby WL.** 2012. The SKP1-like gene family of *Arabidopsis* exhibits a high degree of differential gene expression and gene product interaction during development. *PLoS ONE* **7**, e50984.
- Dinant S, Clark AM, Zhu Y, Vilaine F, Palauqui J-C, Kusiak C, Thompson GA.** 2003. Diversity of the superfamily of phloem lectins (phloem protein 2) in angiosperms. *Plant Physiology* **131**, 114–28.
- Dos Santos C, Delavault P, Letousey P, Thalouarn P.** 2003. Identification by suppression subtractive hybridization and expression analysis of *Arabidopsis thaliana* putative defence genes during *Orobanche ramosa* infection. *Physiological and Molecular Plant Pathology* **62**, 297–303.
- Dumon-Seignovert L, Cariot G, Vuillard L.** 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification* **37**, 203–206.
- Echevarría-Zomeño S, Fernández-Calvino L, Castro-Sanz AB, López JA, Vázquez J, Castellano MM.** 2016. Dissecting the proteome dynamics of the early heat stress response leading to plant survival or death in *Arabidopsis*. *Plant, Cell and Environment* **39**, 1264–1278.
- Edwards K, Johnstone C, Thompson C.** 1991. A simple and rapid method for the preparation of plant DNA for PCR analysis. *Nucleic Acids Research* **19**, 1349–1349.
- Eggermont L, Verstraeten B, Van Damme EJM.** 2017. Genome-wide screening for lectin motifs in *Arabidopsis thaliana*. *The Plant Genome* **10**, 1–17.
- Fan W, Dong X.** 2002. *In vivo* interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis*. *The Plant Cell Online* **14**, 1377–1389.
- Fan W, Zhao M, Li S, Bai X, Li J, Meng H, Mu Z.** 2016. Contrasting transcriptional responses of PYR1/PYL/RCAR ABA receptors to ABA or dehydration stress between maize seedling leaves and roots. *BMC Plant Biology* **16**, 1–14.
- Feilner T, Hultschig C, Lee J, et al.** 2005. High throughput identification of potential *Arabidopsis* mitogen-activated protein kinases substrates. *Molecular and Cellular Proteomics* **4**, 1558–1568.
- Felix G, Duran JD, Volko S, Boller T.** 1999. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant Journal* **18**, 265–276.

- Feng F, Zhou JM.** 2012. Plant-bacterial pathogen interactions mediated by type III effectors. *Current Opinion in Plant Biology* **15**, 469–476.
- Feraru E, Vosolobě S, Feraru MI, Petrášek J, Kleine-Vehn J.** 2012. Evolution and structural diversification of PILS putative auxin carriers in plants. *Frontiers in Plant Science* **3**, 227.
- Fernandez-Calvino L, Faulkner C, Walshaw J, Saalbach G, Bayer E, Benitez-Alfonso Y, Maule A.** 2011. Arabidopsis plasmodesmal proteome. *PLoS ONE* **6**, e18880.
- Fink JL, Hamilton N.** 2007. DomainDraw: a macromolecular feature drawing program. *In Silico Biology* **7**, 145–150.
- Flowers TJ, Colmer TD.** 2008. Salinity tolerance in halophytes. *New Phytologist* **179**, 945–963.
- Forslund K, Henricson A, Hollich V, Sonnhammer ELL.** 2008. Domain tree-based analysis of protein architecture evolution. *Molecular Biology and Evolution* **25**, 254–264.
- Fouquaert E, Peumans WJ, Vandekerckhove TT, Ongenaert M, Van Damme EJM.** 2009. Proteins with an Euonymus lectin-like domain are ubiquitous in Embryophyta. *BMC Plant Biology* **9**, 136.
- Fu ZQ, Dong X.** 2013. Systemic acquired resistance: turning local infection into global defense. *Annual Review of Plant Biology* **64**, 839–863.
- Fu ZQ, Yan S, Saleh A, et al.** 2012. NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants. *Nature* **486**, 228–233.
- Fukao Y.** 2012. Protein-protein interactions in plants. *Plant and Cell Physiology* **53**, 617–625.
- Gagne JM, Downes BP, Shiu S-H, Durski AM, Vierstra RD.** 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11519–11524.
- García AV, Blanvillain-Baufumé S, Huibers R, Wiermer M, Li G, Gobbato E, Rietz S, Parker JE.** 2010. Balanced nuclear and cytoplasmic activities of EDS1 are required for a complete plant innate immune response. *PLoS Pathogens* **6**, e1000970.
- Garcion C, Métraux J.** 2007. Salicylic acid: biosynthesis, metabolism and signal transduction. *Annual Plant Reviews Volume 24: Plant Hormone Signaling*. 229–255.
- Garvey KJ, Saedi MS, Ito J.** 1986. Nucleotide sequence of Bacillus phage ϕ 29 genes 14 and 15: homology of gene 15 with other phage lysozymes. *Nucleic Acids Research* **14**, 10001–10008.
- Gassmann W, Appel HM, Oliver MJ.** 2016. The interface between abiotic and biotic stress responses. *Journal of Experimental Botany* **67**, 2023–2024.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343–345.
- Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C.** 2010. Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818.
- Golldack D, Li C, Mohan H, Probst N.** 2014. Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Frontiers in Plant Science* **5**, 151.
- Goodstein DM, Shu S, Howson R, et al.** 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**, 1178–1186.
- Görlich D.** 1998. Transport into and out of the cell nucleus. *The EMBO Journal* **17**, 2721–2727.
- Gouget A, Senchou V, Govers F, Sanson A, Barre A, Rougé P, Pont-Lezica R, Canut H.** 2006. Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in Arabidopsis. *Plant Physiology* **140**, 81–90.

- Grosdidier A, Zoete V, Michielin O.** 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* **39**, 270–277.
- Gupta B, Huang B.** 2014. Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. *International Journal of Genomics* **2014**, 1–18.
- Gustafsson C, Govindarajan S, Minshull J.** 2004. Codon bias and heterologous protein expression. *Trends in Biotechnology* **22**, 346–353.
- Hacham Y, Koussevitzky S, Kirma M, Amir R.** 2014. Glutathione application affects the transcript profile of genes in *Arabidopsis* seedling. *Journal of Plant Physiology* **171**, 1444–1451.
- Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H.** 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiology* **148**, 993–1003.
- Harrison SJ, Mott EK, Parsley K, Aspinall S, Gray JC, Cottage A.** 2006. A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. *Plant Methods* **2**, 19.
- He X-J, Zhang Z-G, Yan D-Q, Zhang J-S, Chen S-Y.** 2004. A salt-responsive receptor-like kinase gene regulated by the ethylene signaling pathway encodes a plasma membrane serine/threonine kinase. *Theoretical and Applied Genetics* **109**, 377–383.
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J.** 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, 19.
- Hooper CM, Tanz SK, Castleden IR, Vacher MA, Small ID, Millar AH.** 2014. SUBAcon: a consensus algorithm for unifying the subcellular localization data of the *Arabidopsis* proteome. *Bioinformatics* **30**, 3356–3364.
- Houlès Astoul C, Peumans WJ, Van Damme EJM, Barre A, Bourne Y, Rougé P.** 2002. The size, shape and specificity of the sugar-binding site of the jacalin-related lectins is profoundly affected by the proteolytic cleavage of the subunits. *The Biochemical Journal* **367**, 817–824.
- Howe GA, Jander G.** 2008. Plant immunity to insect herbivores. *Annual Review of Plant Biology* **59**, 41–66.
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Grissem W, Zimmermann P.** 2008. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in Bioinformatics* **2008**, 420747.
- Huang P, Ju HW, Min JH, Zhang X, Kim SH, Yang KY, Kim CS.** 2013. Overexpression of L-type lectin-like protein kinase 1 confers pathogen resistance and regulates salinity response in *Arabidopsis thaliana*. *Plant Science* **203–204**, 98–106.
- Huber AE, Bauerle TL.** 2016. Long-distance plant signaling pathways in response to multiple stressors: the gap in knowledge. *Journal of Experimental Botany* **67**, 2063–2079.
- Ingelsson B, Vener A V.** 2012. Phosphoproteomics of *Arabidopsis* chloroplasts reveals involvement of the STN7 kinase in phosphorylation of nucleoid protein PTAC16. *FEBS Letters* **586**, 1265–1271.
- Islam MM, Tani C, Watanabe-Sugimoto M, Uraji M, Jahan MS, Masuda C, Nakamura Y, Mori IC, Murata Y.** 2009. Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in *Arabidopsis* guard cells. *Plant and Cell Physiology* **50**, 1171–1175.

- Iuchi S, Kobayashi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K.** 2001. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *Plant Journal* **27**, 325–333.
- Jakoby MJ, Falkenhan D, Mader MT, Brininstool G, Wischnitzki E, Platz N, Hudson A, Hülskamp M, Larkin J, Schnittger A.** 2008. Transcriptional profiling of mature *Arabidopsis* trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. *Plant Physiology* **148**, 1583–1602.
- Jefferson RA, Kavanagh TA, Bevan MW.** 1987. GUS fusions: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal* **6**, 3901–3907.
- Jia F, Wang C, Huang J, Yang G, Wu C, Zheng C.** 2015. SCF E3 ligase PP2-B11 plays a positive role in response to salt stress in *Arabidopsis*. *Journal of Experimental Botany* **66**, 4683–4697.
- Jones P, Binns D, Chang HY, et al.** 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.
- Jones JDG, Dangl JL.** 2006. The plant immune system. *Nature* **444**, 323–329.
- Käll L, Krogh A, Sonnhammer ELL.** 2004. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**, 1027–1036.
- Kane JF.** 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **6**, 494–500.
- Karimi M, Inzé D, Depicker A.** 2002. GATEWAYTM vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* **7**, 193–195.
- Katagiri F, Thilmony R, He SY.** 2002. The *Arabidopsis thaliana*-*Pseudomonas syringae* interaction. *The Arabidopsis Book* **1**, e0039.
- Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
- Kazan K.** 2015. Diverse roles of jasmonates and ethylene in abiotic stress tolerance. *Trends in Plant Science* **20**, 219–229.
- Kelley L a, Sternberg MJ.** 2015. Partial protein domains: evolutionary insights and bioinformatics challenges. *Genome Biology* **16**, 15–17.
- Kempema LA, Cui X, Holzer FM, Walling LL.** 2006. *Arabidopsis* transcriptome changes in response to phloem-feeding silverleaf whitefly nymphs. Similarities and distinctions in responses to aphids. *Plant Physiology* **143**, 849–865.
- Kerppola TK.** 2008. Bimolecular Fluorescence Complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics* **37**, 465–487.
- Kharenko OA, Choudhary P, Loewen MC.** 2013. Abscisic acid binds to recombinant *Arabidopsis thaliana* G-protein coupled receptor-type G-protein 1 in *Saccharomyces cerevisiae* and *in vitro*. *Plant Physiology and Biochemistry* **68**, 32–36.
- Kim HS, Jung MS, Lee K, Kim KE, Yoo JH, Kim MC, Kim DH, Cho MJ, Chung WS.** 2009. An S-locus receptor-like kinase in plasma membrane interacts with calmodulin in *Arabidopsis*. *FEBS Letters* **583**, 36–42.
- Kipreos ET, Pagano M.** 2000. The F-box protein family. *Genome Biology* **1**, 1–7.
- Kitaoku Y, Fukamizo T, Numata T, Ohnuma T.** 2016. Chitin oligosaccharide binding to the lysin motif of a novel type of chitinase from the multicellular green alga, *Volvox carteri*. *Plant Molecular Biology* **93**, 97–108.

- Kong XY, Kissen R, Bones AM.** 2012. Characterization of recombinant nitrile-specifier proteins (NSPs) of *Arabidopsis thaliana*: dependency on Fe(II) ions and the effect of glucosinolate substrate and reaction conditions. *Phytochemistry* **84**, 7–17.
- 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 α . *Journal of Biological Chemistry* **284**, 478–485.
- Krčková Z, Brouzdová J, Daněk M, Kocourková D, Rainteau D, Ruelland E, Valentová O, Pejchar P, Martinec J.** 2015. Arabidopsis non-specific phospholipase C1: characterization and its involvement in response to heat stress. *Frontiers in Plant Science* **6**, 1–13.
- Krieger E, Koraimann G, Vriend G.** 2002. Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. *Proteins: Structure, Function and Genetics* **47**, 393–402.
- Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G.** 2004. The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. *The Plant Cell* **16**, 3496–3507.
- Kuroda H, Yanagawa Y, Takahashi N, Horii Y, Matsui M.** 2012. A comprehensive analysis of interaction and localization of Arabidopsis SKP1-like (ASK) and F-box (Fbx) proteins. *PLoS ONE* **7**, e50009.
- Kuromori T, Shinozaki K.** 2010. ABA transport factors found in Arabidopsis ABC transporters. *Plant signaling & behavior* **5**, 1124–1126.
- Kus JV, Zaton K, Sarkar R, Cameron RK.** 2002. Age-related resistance in Arabidopsis is a developmentally regulated defense response to *Pseudomonas syringae*. *The Plant Cell* **14**, 479–490.
- Kushalappa AC, Yogendra KN, Karre S.** 2016. Plant innate immune response: qualitative and quantitative resistance. *Critical Reviews in Plant Sciences* **35**, 38–55.
- Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Lalonde S, Ehrhardt DW, Loqué D, Chen J, Rhee SY, Frommer WB.** 2008. Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *The Plant Journal* **53**, 610–635.
- Lamesch P, Berardini TZ, Li D, et al.** 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**, 1202–1210.
- Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH.** 2007. Classical nuclear localization signals: definition, function, and interaction with importin α . *Journal of Biological Chemistry* **282**, 5101–5105.
- Lannoo N, Van Damme EJM.** 2010. Nucleocytoplasmic plant lectins. *Biochimica et Biophysica Acta* **1800**, 190–201.
- Lannoo N, Van Damme EJM.** 2014. Lectin domains at the frontiers of plant defense. *Frontiers in Plant Science* **5**, 397.
- Lannoo N, Peumans WJ, Van Damme EJ.** 2008. Do F-box proteins with a C-terminal domain homologous with the tobacco lectin play a role in protein degradation in plants? *Biochemical Society Transactions* **36**, 843–847.
- Lannoo N, Peumans WJ, Van Pamel E, Alvarez R, Xiong TC, Hause G, Mazars C, Van Damme EJM.** 2006. Localization and *in vitro* binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact *in situ* with high-mannose and complex *N*-glycans. *FEBS Letters* **580**, 6329–6337.

- Lannoo N, Vandenborre G, Miersch O, Smagghe G, Wasternack C, Peumans WJ, Van Damme EJM.** 2007a. The jasmonate-induced expression of the *Nicotiana tabacum* leaf lectin. *Plant and Cell Physiology* **48**, 1207–1218.
- Lannoo N, Vervecken W, Proost P, Rougé P, Van Damme EJM.** 2007b. Expression of the nucleocytoplasmic tobacco lectin in the yeast *Pichia pastoris*. *Protein Expression and Purification* **53**, 275–282.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283–291.
- Lavagi I, Estelle M, Weckwerth W, Beynon J, Bastow RM.** 2012. From bench to bountiful harvests: a road map for the next decade of Arabidopsis research. *The Plant Cell Online* **24**, 2240–2247.
- Lee JR, Boltz KA, Lee SY.** 2014. Molecular chaperone function of *Arabidopsis thaliana* phloem protein 2-A1, encodes a protein similar to phloem lectin. *Biochemical and Biophysical Research Communications* **443**, 18–21.
- Le Gall H, Philippe F, Domon J-M, Gillet F, Pelloux J, Rayon C.** 2015. Cell wall metabolism in response to abiotic stress. *Plants* **4**, 112–166.
- Léon-Kloosterziel KM, Gil MA, Ruijs GJ, Jacobsen SE, Olszewski NE, Schwartz SH, Zeevaart JAD, Koornneef M.** 1996. Isolation and characterization of abscisic acid-deficient Arabidopsis mutants at two new loci. *Plant Journal* **10**, 655–661.
- Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, Rouzé P, Rombauts S.** 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Research* **30**, 325–327.
- Letunic I, Doerks T, Bork P.** 2009. SMART 6: recent updates and new developments. *Nucleic Acids Research* **37**, 229–232.
- Levitt M.** 2009. Nature of the protein universe. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11079–11084.
- Li H, Greene LH.** 2010. Sequence and structural analysis of the chitinase insertion domain reveals two conserved motifs involved in chitin-binding. *PLoS ONE* **5**, e8654.
- Liashkovich I, Meyring A, Oberleithner H, Shahin V.** 2012. Structural organization of the nuclear pore permeability barrier. *Journal of Controlled Release* **160**, 601–608.
- Li L, Yi H.** 2012. Differential expression of Arabidopsis defense-related genes in response to sulfur dioxide. *Chemosphere* **87**, 718–724.
- Lindeberg M, Cunnac S, Collmer A.** 2012. *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends in Microbiology* **20**, 199–208.
- Louis J, Shah J.** 2013. *Arabidopsis thaliana*-*Myzus persicae* interaction: shaping the understanding of plant defense against phloem-feeding aphids. *Frontiers in Plant Science* **4**, 213.
- Lu H, McClung CR, Zhang C.** 2017. Tick tock: circadian regulation of plant innate immunity. *Annual Review of Phytopathology* **55**, 287–311.
- Ludvigsen S, Poulsen FM.** 1992. Secondary structure in solution of barwin from barley seed using 1H nuclear magnetic resonance spectroscopy. *Biochemistry* **31**, 8771–8782.
- Mahajan S, Tuteja N.** 2005. Cold, salinity and drought stresses: an overview. *Archives of Biochemistry and Biophysics* **444**, 139–158.
- Malkov N, Fliegmann J, Rosenberg C, Gascioli V, Timmers ACC, Nurisso A, Cullimore J, Bono J-J.** 2016. Molecular basis of lipo-chitoooligosaccharide recognition by the lysin motif receptor-like kinase LYR3 in legumes. *Biochemical Journal* **473**, 1369–1378.

- Malladi A, Burns JKJ.** 2007. Communication by plant growth regulators in roots and shoots of horticultural crops. *HortScience* **42**, 1113–1117.
- Matsushima R, Fukao Y, Nishimura M, Hara-Nishimura I.** 2004. NA11 gene encodes a basic-helix-loop-helix-type putative transcription factor that regulates the formation of an endoplasmic reticulum-derived structure, the ER body. *The Plant Cell* **16**, 1536–1549.
- Matsushima R, Kondo M, Nishimura M, Hara-Nishimura I.** 2003. A novel ER-derived compartment, the ER body, selectively accumulates a β -glucosidase with an ER-retention signal in *Arabidopsis*. *Plant Journal* **33**, 493–502.
- McElver J, Tzafrir I, Aux G, et al.** 2001. Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* **159**, 1751–1763.
- Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M.** 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**, 662–682.
- Melo F, Feytmans E.** 1998. Assessing protein structures with a non-local atomic interaction energy. *Journal of Molecular Biology* **277**, 1141–1152.
- Mengiste T.** 2012. Plant immunity to necrotrophs. *Annual Review of Phytopathology* **50**, 267–294.
- Mishina TE, Zeier J.** 2007. Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in *Arabidopsis*. *Plant Journal* **50**, 500–513.
- Miura K, Tada Y.** 2014. Regulation of water, salinity, and cold stress responses by salicylic acid. *Frontiers in Plant Science* **5**, 4.
- Monaghan J, Germain H, Weihmann T, Li X.** 2010. Dissecting plant defence signal transduction: modifiers of *snc1* in *Arabidopsis*. *Canadian Journal of Plant Pathology* **32**, 35–42.
- Moore AD, Björklund ÅK, Ekman D, Bornberg-Bauer E, Elofsson A.** 2008. Arrangements in the modular evolution of proteins. *Trends in Biochemical Sciences* **33**, 444–451.
- Moore JW, Loake GJ, Spoel SH.** 2011. Transcription dynamics in plant immunity. *The Plant Cell* **23**, 2809–2820.
- Morris CE, Monteil CL, Berge O.** 2013. The life history of *Pseudomonas syringae*: linking agriculture to earth system processes. *Annual Review of Phytopathology* **51**, 85–104.
- Mukherjee AK, Carp MJ, Zuchman R, Ziv T, Horwitz BA, Gepstein S.** 2010. Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*. *Journal of Proteomics* **73**, 709–720.
- Muthamilarasan M, Prasad M.** 2013. Plant innate immunity: an updated insight into defense mechanism. *Journal of Biosciences* **38**, 433–449.
- Nagano AJ, Fukao Y, Fujiwara M, Nishimura M, Hara-Nishimura I.** 2008. Antagonistic jacalin-related lectins regulate the size of ER body-type β -glucosidase complexes in *Arabidopsis thaliana*. *Plant and Cell Physiology* **49**, 969–980.
- Nagano AJ, Matsushima R, Hara-Nishimura I.** 2005. Activation of an ER-body-localized β -glucosidase via a cytosolic binding partner in damaged tissues of *Arabidopsis thaliana*. *Plant and Cell Physiology* **46**, 1140–1148.
- Nasir A, Kim KM, Caetano-Anollés G.** 2014. Global patterns of protein domain gain and loss in superkingdoms. *PLoS Computational Biology* **10**, e1003452.
- Nitz I, Berkefeld H, Puzio PS, Grundler FMW.** 2001. *Pyk10*, a seedling and root specific gene and promoter from *Arabidopsis thaliana*. *Plant Science* **161**, 337–346.

- Ohnuma T, Numata T, Osawa T, Mizuhara M, Lampela O, Juffer AH, Skriver K, Fukamizo T.** 2011. A class V chitinase from *Arabidopsis thaliana*: gene responses, enzymatic properties, and crystallographic analysis. *Planta* **234**, 123–137.
- Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP.** 2013. Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. *Journal of Experimental Botany* **64**, 445–458.
- Paine PL, Moore LC, Horowitz SB.** 1975. Nuclear envelope permeability. *Nature* **254**, 109–114.
- Paparella C, Savatin DV, Marti L, De Lorenzo G, Ferrari S.** 2014. The *Arabidopsis* LYSIN MOTIF-CONTAINING RECEPTOR-LIKE KINASE3 regulates the cross talk between immunity and abscisic acid responses. *Plant Physiology* **165**, 262–276.
- Parani M, Rudrabhatla S, Myers R, Weirich H, Smith B, Leaman DW, Goldman SL.** 2004. Microarray analysis of nitric oxide responsive transcripts in *Arabidopsis*. *Plant Biotechnology Journal* **2**, 359–366.
- Passarinho PA, Van Hengel AJ, Fransz PF, De Vries SC.** 2001. Expression pattern of the *Arabidopsis thaliana* AtEP3/AtchitIV endochitinase gene. *Planta* **212**, 556–567.
- Pemberton LF, Paschal BM.** 2005. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187–198.
- Pesch M, Hülkamp M.** 2004. Creating a two-dimensional pattern *de novo* during *Arabidopsis* trichome and root hair initiation. *Current Opinion in Genetics and Development* **14**, 422–427.
- Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.** 2004. UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612.
- Petutschnig EK, Jones AME, Serazetdinova L, Lipka U, Lipka V.** 2010. The Lysin Motif Receptor-like Kinase (LysM-RLK) CERK1 is a major chitin-binding protein in *Arabidopsis thaliana* and subject to chitin-induced phosphorylation. *Journal of Biological Chemistry* **285**, 28902–28911.
- Peumans WJ, Van Damme EJM.** 1995. Lectins as plant defense proteins. *Plant Physiology* **109**, 347–352.
- Peumans WJ, Hause B, Van Damme EJM.** 2000. The galactose-binding and mannose-binding jacalin-related lectins are located in different sub-cellular compartments. *FEBS Letters* **477**, 186–192.
- Pfaffl MW, Horgan GW, Dempfle L.** 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* **30**, e36.
- Pieterse CMJ, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SCM.** 2012. Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology* **28**, 489–521.
- Pieterse CMJ, van Wees SC, Hoffland E, van Pelt JA, van Loon LC.** 1996. Systemic resistance in *Arabidopsis* induced by biocontrol bacteria is independent of salicylic acid accumulation and pathogenesis-related gene expression. *The Plant Cell* **8**, 1225–1237.
- Price NPJ, Momany FA, Schnupf U, Naumann TA.** 2015. Structure and disulfide bonding pattern of the hevein-like peptide domains from plant class IV chitinases. *Physiological and Molecular Plant Pathology* **89**, 25–30.
- Qu A-L, Ding Y-F, Jiang Q, Zhu C.** 2013. Molecular mechanisms of the plant heat stress response. *Biochemical and Biophysical Research Communications* **432**, 203–207.

- Rao VS, Srinivas K, Sujini GN, Kumar GNS.** 2014. Protein-protein interaction detection: methods and analysis. *International Journal of Proteomics* **2014**, 1–12.
- Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J.** 2013. Transcriptome responses to combinations of stresses in *Arabidopsis*. *Plant Physiology* **161**, 1783–1794.
- Raval S, Gowda SB, Singh DD, Chandra NR.** 2004. A database analysis of jacalin-like lectins: sequence-structure-function relationships. *Glycobiology* **14**, 1247–1263.
- Reigh DL, Wender SH, Smith EC.** 1973. Scopoletin: a substrate for an isoperoxidase from *Nicotiana tabacum* tissue culture W-38. *Phytochemistry* **12**, 1265–1268.
- Rentzsch R, Orengo CA.** 2013. Protein function prediction using domain families. *BMC Bioinformatics* **14**, S5.
- Reuber TL, Ausubel FM.** 1996. Isolation of *Arabidopsis* genes that differentiate between resistance responses mediated by the RPS2 and RPM1 disease resistance genes. *The Plant Cell* **8**, 241–249.
- Revers F, Guiraud T, Houvenaghel M-C, Mauduit T, Le Gall O, Candresse T.** 2003. Multiple resistance phenotypes to lettuce mosaic virus among *Arabidopsis thaliana* accessions. *Molecular Plant-Microbe Interactions* **16**, 608–616.
- Rhee SY, Beavis W, Berardini TZ, et al.** 2003. The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Research* **31**, 224–228.
- Risseuw EP, Daskalchuk TE, Banks TW, Liu E, Cotelesage J, Hellmann H, Estelle M, Somers DE, Crosby WL.** 2003. Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant Journal* **34**, 753–767.
- Roden LC, Ingle RA.** 2009. Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant-pathogen interactions. *The Plant Cell Online* **21**, 2546–2552.
- Rosano GL, Ceccarelli EA.** 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology* **5**, 172.
- Roychoudhury A, Paul S, Basu S.** 2013. Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. *Plant Cell Reports* **32**, 985–1006.
- Sabol P, Kulich I, Žárský V.** 2017. RIN4 recruits the exocyst subunit EXO70B1 to the plasma membrane. *Journal of Experimental Botany* **68**, 3253–3265.
- Sahdev S, Khatyar SK, Saini KS.** 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry* **307**, 249–264.
- Sakamoto H, Matsuda O, Iba K.** 2008. ITN1, a novel gene encoding an ankyrin-repeat protein that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance in *Arabidopsis thaliana*. *Plant Journal* **56**, 411–422.
- Salas-Muñoz S, Gómez-Anduro G, Delgado-Sánchez P, Rodríguez-Kessler M, Jiménez-Bremont JF.** 2012. The *Opuntia streptacantha* OpsHSP18 gene confers salt and osmotic stress tolerance in *Arabidopsis thaliana*. *International Journal of Molecular Sciences* **13**, 10154–10175.
- Salas-Muñoz S, Rodríguez-Hernández AA, Ortega-Amaro MA, Salazar-Badillo FB, Jiménez-Bremont JF.** 2016. *Arabidopsis AtDJA3* null mutant shows increased sensitivity to abscisic acid, salt, and osmotic stress in germination and post-germination stages. *Frontiers in Plant Science* **7**, 1–11.

- Sanabria N, Goring D, Nürnberger T, Dubery I.** 2008. Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. *New Phytologist* **178**, 503–514.
- Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M.** 1996. A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a β -prism fold. *Nature Structural Biology* **3**, 596–603.
- Schellmann S, Hülskamp M, Uhrig J.** 2007. Epidermal pattern formation in the root and shoot of *Arabidopsis*. *Biochemical Society Transactions* **35**, 146–148.
- Schindelin J, Arganda-Carreras I, Frise E, et al.** 2012. Fiji: an open source platform for biological image analysis. *Nature Methods* **9**, 676–682.
- Schouppe D, Ghesquière B, Menschaert G, De Vos WH, Bourque S, Trooskens G, Proost P, Gevaert K, Van Damme EJM.** 2011. Interaction of the tobacco lectin with histone proteins. *Plant Physiology* **155**, 1091–1102.
- Schouppe D, Rougé P, Lasanajak Y, Barre A, Smith DF, Proost P, Van Damme EJM.** 2010. Mutational analysis of the carbohydrate binding activity of the tobacco lectin. *Glycoconjugate Journal* **27**, 613–623.
- Seo PJ, Lee AK, Xiang F, Park CM.** 2008. Molecular and functional profiling of *Arabidopsis* pathogenesis-related genes: insights into their roles in salt response of seed germination. *Plant and Cell Physiology* **49**, 334–344.
- Sewelam N, Oshima Y, Mitsuda N, Ohme-Takagi M.** 2014. A step towards understanding plant responses to multiple environmental stresses: a genome-wide study. *Plant, Cell and Environment* **37**, 2024–2035.
- Shahidi-Noghabi S, Van Damme EJM, Smagghe G.** 2009. Expression of *Sambucus nigra* agglutinin (SNA-I) from elderberry bark in transgenic tobacco plants results in enhanced resistance to different insect species. *Transgenic Research* **18**, 249–259.
- Shan X, Yan J, Xie D.** 2012. Comparison of phytohormone signaling mechanisms. *Current Opinion in Plant Biology* **15**, 84–91.
- Sherameti I, Venus Y, Drzewiecki C, Tripathi S, Dan VM, Nitz I, Varma A, Grundler FM, Oelmüller R.** 2008. PYK10, a β -glucosidase located in the endoplasmic reticulum, is crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus *Piriformospora indica*. *The Plant Journal* **54**, 428–439.
- Shi H, Lee B, Wu S-J, Zhu J-K.** 2003. Overexpression of a plasma membrane Na⁺/H⁺ antiporter gene improves salt tolerance in *Arabidopsis thaliana*. *Nature Biotechnology* **21**, 81–85.
- Sievers F, Wilm A, Dineen D, et al.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539.
- Simpson PJ, Jamieson SJ, Abou-Hachem M, Nordberg Karlsson E, Gilbert HJ, Holst O, Williamson MP.** 2002. The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. *Biochemistry* **41**, 5712–5719.
- Singh P, Kuo Y-C, Mishra S, et al.** 2012. The lectin receptor kinase-VI.2 is required for priming and positively regulates *Arabidopsis* pattern-triggered immunity. *Plant Cell* **24**, 1256–1270.
- Sivasubramanian R, Mukhi N, Kaur J.** 2015. *Arabidopsis thaliana*: a model for plant research. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*. 1–26.
- Skaar JR, Pagan JK, Pagano M.** 2013. Mechanisms and function of substrate recruitment by F-box proteins. *Nature Reviews Molecular Cell Biology* **14**, 369–381.

- Smedley D, Haider S, Durinck S, et al.** 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research* **43**, 589–598.
- Smith JL, De Moraes CM, Mescher MC.** 2009. Jasmonate- and salicylate-mediated plant defense responses to insect herbivores, pathogens and parasitic plants. *Pest Management Science* **65**, 497–503.
- Soderholm JF, Bird SL, Kalab P, Sampathkumar Y, Hasegawa K, Uehara-Bingen M, Weis K, Heald R.** 2011. Importazole, a small molecule inhibitor of the transport receptor importin- β . *American Chemical Society Chemical Biology* **6**, 700–708.
- Song M, Xu W, Xiang Y, Jia H, Zhang L, Ma Z.** 2014. Association of jacalin-related lectins with wheat responses to stresses revealed by transcriptional profiling. *Plant Molecular Biology* **84**, 95–110.
- Sparkes IA, Runions J, Kearns A, Hawes C.** 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols* **1**, 2019–2025.
- Spoel SH, Mou Z, Tada Y, Spivey NW, Genschik P, Dong X.** 2009. Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity. *Cell* **137**, 860–872.
- Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
- Stamatakis A, Hoover P, Rougemont J.** 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* **57**, 758–771.
- Stefanowicz K.** 2015. Involvement of a carbohydrate-binding F-box-Nictaba protein from *Arabidopsis thaliana* in plant stress responses. PhD thesis, Ghent University.
- Stefanowicz K, Lannoo N, Proost P, Van Damme EJM.** 2012. Arabidopsis F-box protein containing a Nictaba-related lectin domain interacts with *N*-acetylglucosamine structures. *FEBS Open Bio* **2**, 151–158.
- Stefanowicz K, Lannoo N, Zhao Y, Eggermont L, Van Hove J, Al Atalah B, Van Damme EJM.** 2016. Glycan-binding F-box protein from *Arabidopsis thaliana* protects plants from *Pseudomonas syringae* infection. *BMC Plant Biology* **16**, 213.
- Suh EK, Gumbiner BM.** 2003. Translocation of β -catenin into the nucleus independent of interactions with FG-rich nucleoporins. *Experimental Cell Research* **290**, 447–456.
- Surico G.** 2013. The concepts of plant pathogenicity, virulence/avirulence and effector proteins by a teacher of plant pathology. *Phytopathologia Mediterranea* **52**, 399–417.
- Suzuki N, Bassil E, Hamilton JS, et al.** 2016. ABA is required for plant acclimation to a combination of salt and heat stress. *PLoS ONE* **11**, e0147625.
- Swarbreck D, Wilks C, Lamesch P, et al.** 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* **36**, 1009–1014.
- Tada Y, Spoel SH, Pajerowska-Mukhtar K, Mou Z, Song J, Wang C, Zuo J, Dong X.** 2008. Plant immunity requires conformational changes of NPR1 via S-nitrosylation and thioredoxins. *Science* **321**, 952–956.
- Takahashi N, Kuroda H, Kuromori T, Hirayama T, Seki M, Shinozaki K, Shimada H, Matsui M.** 2004. Expression and interaction analysis of Arabidopsis Skp1-related genes. *Plant and Cell Physiology* **45**, 83–91.
- Thangstad OP, Gilde B, Chadchawan S, Seem M, Husebye H, Bradley D, Bones AM.** 2004. Cell specific, cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant Molecular Biology* **54**, 597–611.

- The Arabidopsis Genome Initiative.** 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Thomma BP, Eggermont K, Tierens KF, Broekaert WF.** 1999. Requirement of functional ethylene-insensitive 2 gene for efficient resistance of Arabidopsis to infection by *Botrytis cinerea*. *Plant Physiology* **121**, 1093–1102.
- Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, Yamaguchi-Shinozaki K.** 2007. Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20623–20628.
- Tsuda K, Katagiri F.** 2010. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology* **13**, 459–465.
- Tsuzuki T, Takahashi K, Inoue S ichiro, Okigaki Y, Tomiyama M, Hossain MA, Shimazaki K ichiro, Murata Y, Kinoshita T.** 2011. Mg-chelatase H subunit affects ABA signaling in stomatal guard cells, but is not an ABA receptor in *Arabidopsis thaliana*. *Journal of Plant Research* **124**, 527–538.
- Vaid N, Macovei A, Tuteja N.** 2013. Knights in action: lectin receptor-like kinases in plant development and stress responses. *Molecular Plant* **6**, 1405–1418.
- Valenzuela CE, Acevedo-Acevedo O, Miranda GS, Vergara-Barros P, Holuigue L, Figueroa CR, Figueroa PM.** 2016. Salt stress response triggers activation of the jasmonate signaling pathway leading to inhibition of cell elongation in Arabidopsis primary root. *Journal of Experimental Botany* **67**, 4209–4220.
- Van Damme EJM.** 2014. History of plant lectin research. *Methods in Molecular Biology* **1200**, 3–13.
- Van Damme EJM, Barre A, Rougé P, Peumans WJ.** 2004. Cytoplasmic/nuclear plant lectins: a new story. *Trends in Plant Science* **9**, 484–489.
- Van Damme EJM, Hao Q, Chen Y, Barre A, Vandenbussche F, Desmyter S, Rouge P, Peumans WJ.** 2001. Ribosome-inactivating proteins: a family of plant proteins that do more than inactivate ribosomes. *Critical Reviews in Plant Sciences* **20**, 395–466.
- Van Damme EJM, Lannoo N, Peumans WJ.** 2008. Plant lectins. *Advances in Botanical Research* **48**, 107–209.
- Van Damme EJM, Peumans WJ, Barre A, Rougé P.** 1998. Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Reviews in Plant Sciences* **17**, 575–692.
- Vandenborre G, Van Damme EJM, Smagghe G.** 2009. *Nicotiana tabacum* agglutinin expression in response to different biotic challengers. *Arthropod-Plant Interactions* **3**, 193–202.
- Vandenborre G, Groten K, Smagghe G, Lannoo N, Baldwin IT, Van Damme EJM.** 2010. *Nicotiana tabacum* agglutinin is active against Lepidopteran pest insects. *Journal of Experimental Botany* **61**, 1003–1014.
- Vandenborre G, Smagghe G, Van Damme EJM.** 2011. Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* **72**, 1538–1550.
- Van Holle S.** 2016. Nictaba-like genes from *Glycine max* are part of the plant stress response. PhD thesis, Ghent University.
- Van Holle S, Rougé P, Van Damme EJM.** 2017a. Evolution and structural diversification of Nictaba-like lectin genes in food crops with a focus on soybean (*Glycine max*). *Annals of Botany* **119**, 901–914.

- Van Holle S, De Schutter K, Eggermont L, Tsaneva M, Dang L, Van Damme EJM.** 2017b. Comparative study of lectin domains in model species: new insights into evolutionary dynamics. *International Journal of Molecular Sciences* **18**, 1136.
- Van Holle S, Smaghe G, Van Damme EJM.** 2016. Overexpression of *Nictaba*-like lectin genes from *Glycine max* confers tolerance toward *Pseudomonas syringae* infection, aphid infestation and salt stress in transgenic Arabidopsis plants. *Frontiers in Plant Science* **7**, 1590.
- Van Hove J, Fouquaert E, Smith DF, Proost P, Van Damme EJM.** 2011. Lectin activity of the nucleocytoplasmic EUL protein from *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications* **414**, 101–105.
- Van Hove J, De Jaeger G, De Winne N, Guisez Y, Van Damme EJM.** 2015. The Arabidopsis lectin EULS3 is involved in stomatal closure. *Plant Science* **238**, 312–322.
- Van Hove J, Stefanowicz K, De Schutter K, Eggermont L, Lannoo N, Al Atalah B, Van Damme EJM.** 2014. Transcriptional profiling of the lectin ArathEULS3 from *Arabidopsis thaliana* toward abiotic stresses. *Journal of Plant Physiology* **171**, 1763–1773.
- van Loon LC, Rep M, Pieterse CMJ.** 2006. Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* **44**, 135–162.
- Verslues PE.** 2016. ABA and cytokinins: challenge and opportunity for plant stress research. *Plant Molecular Biology* **91**, 629–640.
- Vishwakarma K, Upadhyay N, Kumar N, et al.** 2017. Abscisic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects. *Frontiers in Plant Science* **8**, 161.
- Voorrips RE.** 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* **93**, 77–78.
- Vos IA, Pieterse CMJ, Van Wees SCM.** 2013. Costs and benefits of hormone-regulated plant defences. *Plant Pathology* **62**, 43–55.
- Wan J, Patel A, Mathieu M, Kim S-Y, Xu D, Stacey G.** 2008. A lectin receptor-like kinase is required for pollen development in Arabidopsis. *Plant Molecular Biology* **67**, 469–482.
- Wan J, Tanaka K, Zhang X-C, Son GH, Brechenmacher L, Nguyen THN, Stacey G.** 2012. LYK4, a lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in Arabidopsis. *Plant Physiology* **160**, 396–406.
- Wang R, Brattain MG.** 2007. The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. *FEBS Letters* **581**, 3164–3170.
- Wang W, Vinocur B, Altman A.** 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**, 1–14.
- Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB.** 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiology* **144**, 32–42.
- Willmann R, Lajunen HM, Erbs G, et al.** 2011. Arabidopsis lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proceedings of the National Academy of Sciences* **108**, 19824–19829.
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ.** 2007. An ‘electronic fluorescent pictograph’ browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2**, e718.
- Wirthmueller L, Maqbool A, Banfield MJ.** 2013. On the front line: structural insights into plant-pathogen interactions. *Nature Reviews Microbiology* **11**, 761–776.

- Wittstock U, Burow M.** 2010. Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *The Arabidopsis Book* **8**, e0134.
- Wu J, Luo X, Guo H, Xiao J, Tian Y.** 2006. Transgenic cotton, expressing *Amaranthus caudatus* agglutinin, confers enhanced resistance to aphids. *Plant Breeding* **125**, 390–394.
- Wu L, Zu X, Zhang H, Wu L, Xi Z, Chen Y.** 2015. Overexpression of ZmMAPK1 enhances drought and heat stress in transgenic *Arabidopsis thaliana*. *Plant Molecular Biology* **88**, 429–443.
- Xiao J, Li C, Xu S, Xing L, Xu Y, Chong K.** 2015. JACALIN-LECTIN LIKE1 regulates the nuclear accumulation of GLYCINE-RICH RNA-BINDING PROTEIN7, influencing the RNA processing of FLOWERING LOCUS C antisense transcripts and flowering time in Arabidopsis. *Plant Physiology* **169**, 2102–2117.
- Xin Z, Wang A, Yang G, Gao P, Zheng Z-L.** 2009. The Arabidopsis A4 subfamily of lectin receptor kinases negatively regulates abscisic acid response in seed germination. *Plant Physiology* **149**, 434–444.
- Xin Y, Xiangrong Z, Mingju Z, Wenchao G, Yingchuan T, Qizhong X, Jiahe W.** 2011. Transgenic potato overexpressing the *Amaranthus caudatus* agglutinin gene to confer aphid resistance. *Crop Science* **51**, 2119–2124.
- Xiong L, Zhu J-K.** 2002. Molecular and genetic aspects of plant responses to osmotic stress. *Plant, Cell and Environment* **25**, 131–139.
- Xu Z, Escamilla-Treviño L, Zeng L, et al.** 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Molecular Biology* **55**, 343–367.
- Xu K, Liu J, Fan M, Xin W, Hu Y, Xu C.** 2012. A genome-wide transcriptome profiling reveals the early molecular events during callus initiation in Arabidopsis multiple organs. *Genomics* **100**, 116–124.
- Xu E, Vaahtera L, Brosché M.** 2015. Roles of defense hormones in the regulation of ozone-induced changes in gene expression and cell death. *Molecular Plant* **8**, 1776–1794.
- Xue J, Jørgensen M, Pihlgren U, Rask L.** 1995. The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. *Plant Molecular Biology* **27**, 911–922.
- Yamaguchi Y, Huffaker A, Bryan AC, Tax FE, Ryan CA.** 2010. PEPR2 is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in Arabidopsis. *The Plant Cell* **22**, 508–522.
- Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L, Grotewold E.** 2011. AGRIS: the Arabidopsis gene regulatory information server, an update. *Nucleic Acids Research* **39**, 1118–1122.
- Younis A, Siddique MI, Kim CK, Lim KB.** 2014. RNA interference (RNAi) induced gene silencing: a promising approach of hi-tech plant breeding. *International Journal of Biological Sciences* **10**, 1150–1158.
- Zaltsman A, Krichevsky A, Loyter A, Citovsky V.** 2010. Agrobacterium induces expression of a host F-box protein required for tumorigenicity. *Cell Host and Microbe* **7**, 197–209.
- Zhang M, Chekan JR, Dodd D, Hong P-Y, Radlinski L, Revindran V, Nair SK, Mackie RI, Cann I.** 2014. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3708–3717.
- Zhang Y, Gao P, Yuan JS.** 2010a. Plant protein-protein interaction network and interactome. *Current Genomics* **11**, 40–46.
- Zhang J, Li W, Xiang T, et al.** 2010b. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host and Microbe* **7**, 290–301.

- Zhang Z-S, Liu M-J, Gao H-Y, Jin L-Q, Li Y-T, Li Q-M, Ai X-Z.** 2015. Water status related root-to-shoot communication regulates the chilling tolerance of shoot in cucumber (*Cucumis sativus* L.) plants. *Scientific Reports* **5**, 13094.
- Zhang Z, Ober JA, Kliebenstein DJ.** 2006. The gene controlling the quantitative trait locus EPITHIOSPECIFIER MODIFIER1 alters glucosinolate hydrolysis and insect resistance in Arabidopsis. *The Plant Cell* **18**, 1524–1536.
- Zhang C, Shi H, Chen L, et al.** 2011. Harpin-induced expression and transgenic overexpression of the phloem protein gene *AtPP2-A1* in Arabidopsis repress phloem feeding of the green peach aphid *Myzus persicae*. *BMC Plant Biology* **11**, 11.
- Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, Cannon SB, Stacey G.** 2007. Molecular evolution of lysin motif-type receptor-like kinases in plants. *Plant Physiology* **144**, 623–636.
- Zhang C, Xie Q, Anderson RG, et al.** 2013. Crosstalk between the circadian clock and innate immunity in Arabidopsis. *PLoS Pathogens* **9**, e1003370.
- Zhu J.** 2007. Plant salt stress. *Encyclopedia of Life Sciences*, 1–3.
- Zhu QH, Stephen S, Kazan K, Jin G, Fan L, Taylor J, Dennis ES, Helliwell CA, Wang MB.** 2013. Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in Arabidopsis using RNA-seq. *Gene* **512**, 259–266.
- Ziemienowicz A, Haasen D, Staiger D, Merkle T.** 2003. Arabidopsis transportin1 is the nuclear import receptor for the circadian clock-regulated RNA-binding protein AtGRP7. *Plant Molecular Biology* **53**, 201–212.
- Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, Boller T, Felix G.** 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* **125**, 749–760.
- Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G, Boller T.** 2004. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature* **428**, 764–767.

Acknowledgements

Oef, het is zover, het einde is in zicht! Dit doctoraat maakte de afgelopen 6 jaar een deel uit van mijn leven. Aan deze jaren hou ik, zowel van op als naast het werk, leuke en minder leuke herinneringen over. Gedurende deze 6 jaar zijn mijn ideeën over het onderzoek en het leven, mede door deze ervaringen, gegroeid. Heel wat mensen rondom mij, zowel op als naast het werk, maakten deel uit van deze ervaringen. Daarvoor wil ik hen hier, aan het einde van deze 6 jaar, dan ook bedanken.

Eerst en vooral, Prof. Els Van Damme, mijn promotor. **Els**, zonder jou was dit doctoraat nooit tot stand gekomen. Je gaf me in 2011 de kans om binnen de Glyco groep als assistent aan de slag te gaan. Tot op de dag van vandaag, ben ik je hiervoor dankbaar en ben ik ook heel blij dat ik deze kans toen gegrepen heb. Gedurende de verschillende jaren stond jouw deur dan ook steeds open, bedankt hiervoor! Het laatste half jaar was ik minder in Gent, maar zaten we toch geregeld samen rond de verschillende hoofdstukken van mijn doctoraat, die je meermaals kritisch hebt nagelezen. Bedankt, om er samen met mij, een boek van te maken waarop ik trots kan zijn!

Ik zou ook graag alle leden van de examencommissie willen bedanken: **Prof. dr. ir. Mieke Uyttendaele, Prof. dr. ir. Kris Audenaert, Prof. dr. ir. Tina Kyndt, Prof. dr. Filip Vandebussche en Prof. dr. Filip Rolland**. Bedankt om tijd te maken om mijn werk in detail te lezen en suggesties te geven die het alleen maar beter maakten!

Of course I would like to thank all members and ex members of the Glyco team. Thank you to create a nice work atmosphere and for all the funny conversations in the office or the labs, moreover thanks for your support whenever necessary! **Sofie**, toen ik in het Glyco team terecht kwam, was jij al een maand begonnen als student bij Annelies. Na jouw jaar als thesisstudent, bleef je bij ons voor een doctoraat. Ik vind het super dat er toch iemand is van de glyco's waarmee ik mijn volledige weg van het doctoraat heb mogen afleggen. Ook al werk je nu als post-doc niet meer op Nictaba, wij vormen nog steeds het Nictaba-team. Als Nictaba-maatjes hebben we samen hard gewerkt, maar ook samen heel veel gelachen. We zorgden ook samen dikwijls voor de algemene organisatie van het labo onder het motto 'samen sterk'. Bij een verjaardag, een doctoraat of een afscheid probeerden we steeds samen te zorgen voor een leuke verrassing of cadeau. Ook met Kerst en Pasen, zorgden we voor de Kerst- en Paasfeer. Bedankt voor de enorme steun, de vele babbels, de uitstapjes en nog zoveel meer... **Jeroen**, de sterke Glyco man. Jij kwam exact een jaar later bij het Glyco team dan Sofie, eerst ook als thesisstudent van Annelies en daarna voor een doctoraat. Op een gegeven moment werden wij trein buddies (behalve 's ochtends :p) en zelf iets later ook eiland buddies. Als ik mij niet vergis, vond jij het spreekwoord 'een Loreke doen' uit en je was er dan ook altijd graag bij wanneer er een zot moment te beleven viel in het bureau of labo. Naast alle zottigheden, konden we ook steeds bij elkaar terecht voor allerlei soorten (soms diepzinnige) gesprekken en kon ik ook steeds op jou rekenen. Bedankt voor het

luisterend oor, de grapjes, de uitstapjes en nog zoveel meer... En uiteraard succes met het verderzetten van mijn werk als assistent voor de studenten. Het zal druk worden, maar ik twijfel er niet aan dat jij dit graag en goed zult doen! **Gosia**, my Glyco hipster friend, it all started during your internship with Sofie and me in the summer of 2015. I know Sofie and me made you work hard, but eventually I think you're glad we learned you all this stuff and you could stay in the Glyco team to start your PhD. I really enjoyed our many lunch moments together and our many West-Flemish-English mixed talks. Of course I also enjoyed the moments with you, Thomas and others at your place and want to thank you and Thomas for your hospitality! I'm very happy you already finished one year of Dutch classes and enjoyed the exercises with you. I hope one day we can speak fluent West-Flemish with each other, that would be awesome. To start this, I will end my text for you in Dutch. Bedankt voor jouw eerlijkheid, de vele grappige momenten, de serieuze babbels en nog zoveel meer... **Isabel**, wat ik mij het best herinner en ook het langst zal onthouden, zijn onze 'plant flow' momentjes. Deze momenten begonnen meestal met een goed gesprek of een leuke babbel tijdens het zaaien met de radio op de achtergrond. Naargelang de uren voorbij tikten en het zaad maar langzaam minderde, zetten we de radio luider en zongen we luidkeels mee. Zo was het saaie zaaiwerk aan de 'plant flow' steeds een feest! Bedankt voor de vele hulp in het labo en jouw, door niemand gedurfde, fantastische grappen! **Mariya**, you are next to finish your PhD. Your project is not the easiest one, but you are a tough woman and I'm sure you will bring your struggle with the rice EULs to a good end! Thanks for the many facts about Bulgaria and for the nice postcards and sweets each time you went on a trip. **Subbu**, although you are a more silent type, we had some nice and funny moments together, thank you! Moreover, I wish you all the best in the future and I hope one time you will be able to start your company, good luck. **Pieter en Jeroen**, peppi en kokki van het Glyco team, jullie vormen een goed team en zullen, samen met Sinem, waarschijnlijk nog het langst binnen het Glyco team werken. Succes met jullie doctoraat en zorg goed voor nieuwkomertjes binnen het team. Bedankt Pieter voor de vele grappige discussies met Jeroen DZ en bedankt Jeroen om een beetje de radio van onze bureau te zijn! **Sinem**, also you will spend some more years in the Glyco team, take care of the new PhD students in the team and good luck with your PhD. Thank you for the nice time and I will still think about you when I'm using my 'hoelahoep'! **Kristof**, je bent dan wel weg uit het Glyco bureau, in de Glyco meetings ben je nog steeds aanwezig. Ook tijdens de lunch pauzes zien we jou steeds, meestal met het picknick mandje die je deelt met Isabel. Bedankt voor jouw wetenschappelijk advies wanneer nodig en de grappige momenten in het bureau. **Freja**, het zonnestraaltje, ook al zagen we elkaar niet zo heel veel, het was altijd fijn om een snelle babbel met jou te slaan, bedankt daarvoor! **Ying, Weidong and Chen**, it was always nice to meet you in the corridors and during the several Glyco meetings.

Several Glyco members who worked in the team when I started, already left our Glyco team. Of course I want to thank them also! **Annelies**, mijn West-Vlaamse maatje, het was super om bij jou in het Nictaba-team terecht te komen. Bedankt om mij wegwijs te maken in het labo en zeker ook voor onze vele avondjes uit in Gent, fun verzekerd! **Nausicaä**, onze Glyco

mama, ik weet nog heel goed dat jij mij de bomen door het bos liet zien wanneer er gedacht moest worden aan een doctoraat neerschrijven. Bedankt voor de steun en natuurlijk bedankt voor de vele lachmomentjes en leuke uitjes! **Liuyi**, because of you the cold room work was filled with nice music, thank you! **Kirsten**, ik herinner mij jou als iemand die de alledaagse dingen steeds op een grappige manier kon vertellen, zeker over jouw kindjes, het was super om een aantal jaar met jou in het Glyco team te kunnen werken, bedankt! **Bassam**, the 'think positive' Glyco member, it's a pity we cannot stay in contact with each other, I will never forget how you always made me smile even in 'bad' and busy moments, thanks! **Shang**, I remember your chaos in the labs, but you are such a nice person, always telling funny stories, thank you! **Jonas**, mijn vroegere eiland buddy, bedankt voor de vele grappen en jouw wetenschappelijk advies. Het was altijd fijn om met jou samen te werken voor het vak Eiwitchemie! And of course also thanks to **Karolina, Elke, Tomasz**, my students **Jana, Oriana, Bruno, Jonathan** and **Karolina** for the lovely memories and lots of funny moments! Thanks to all GLYCO members for the nice times!

Next to the Glyco team, I would like to thank all the people from the department of Molecular Biotechnology for the nice lunch breaks, the funny moments, the help,... **Lander**, jij was er van in het begin en bent er nu nog steeds, super bedankt voor de zes TOP jaren als collega AAP! We hebben het samen soms goed druk gehad, maar konden het gelukkig altijd samen fiksen. Ook bedankt om mij zoveel aan te leren in mijn eerste jaar, ik kon mijn geen betere 'leraar' wensen! Bij jou kon ik ook steeds met alles terecht, een grote dankjewel voor alle hulp, oppeppertjes, grapjes, babbeltjes en nog zoveel meer... **Sofie, Fien en Aïsj**a, ons topteam in het secretariaat waaraan je alles kon vragen, bedankt voor de vele babbeltjes, jullie glimlach, de steun,... **Geert**, onze technische mobi man, bedankt om mij steeds uit de nood te helpen met de qPCR robot, de microscoop en alle andere toestellen die mij wel eens op de zenuwen gewerkt hebben. **Lien**, bedankt voor de vele gezellige treinritten en babbeltjes in de gang! **Tom** (of is het Tommie ;)), onze Limburger, ook al zagen we elkaar niet zoveel, het was steeds tof om een babbeltje te slaan. Bedankt ook voor het organiseren van de jaarlijkse 'Monkey day', top! **Ruben, Henok, Richard, Diana, Jonas, Pieter**, ... thanks! **Sarah**, de laatste jaren geen mobi meer, maar een super collega en vriendin. Bedankt voor de vele knuffels, babbeltjes, gezellige avondjes, ... Ook dankjewel aan de andere immuno's, **Annelien en Stefanie**, voor de vele gezellige momentjes in het bureau, het was super om het bureau met jullie te delen. Voor de mensen die hun naam hier niet zien staan, sorry, ik heb echt te weinig plaats en vond het zeker ook fijn om, samen met jullie, deel uit te maken van dezelfde vakgroep!

Naast de vele collega's die er waren voor mij op de werkvloer en voor zes aangename jaren zorgden, hebben natuurlijk ook vele vrienden mij gesteund tijdens mijn doctoraat en gezorgd voor de vele ontspannende avonden. **Iris, Fre, Lisa** en **Sanne**, ik ben nu reeds een jaartje metie van jullie kleinste spruit Sanne en daar ben ik heel gelukkig mee. Bedankt voor de vele gezellige uitstapjes, avondjes en zeker ook jullie steun wanneer het nodig was. **Kimberly** en **Fran**, mijn muisjesvriendinnen, ik weet niet wie er nu 'pepe' heeft? ;) Bedankt

voor de vele super dineetjes en avondjes uit in Kortrijk. Bedankt ook voor jullie onvoorwaardelijke steun wanneer ik een moeilijke periode had! **Charlotte, Justine, Helene, Anne-Laure** en **Delphine**, bedankt voor de vele gezellige en ontspannende etentjes, spelletjes-avonden, ... Bedankt ook voor jullie motiverende woorden wanneer de motivatie eventjes ver te zoeken was! **Suzanne**, bedankt voor de vele toffe babbels en het aangenaam en motiverend gezelschap op de trein (naast de kulak reünie momentjes). **Eline**, mijn Engeline, bedankt voor de motiverende babbels en kaartjes! De **Spessers**, de **vrienden van het muziek** en de **vrienden van Bjorn**, teveel mensen om bij naam te noemen, bedankt voor de vele gezellige namiddagen, avonden, zondagvoormiddagen, chiroweekendjes en kampen!

Bjorn, mijn liefje, zoals we zoveel momenten gekend hebben het laatste half jaar, zit ik nu aan mijn computer dit dankwoord te schrijven en zit jij naast mij aan jouw computer te programmeren. Ik denk dat er teveel dingen zijn waarvoor ik jou wil bedanken, maar ik probeer er hier toch de belangrijkste aan te halen. Bedankt voor jouw liefde, steun, advies, hulp, eerlijkheid, vele telefoontjes en natuurlijk knuffels! Je hebt slechts het laatste jaar van mijn doctoraat van dichtbij meegemaakt, maar dan wel het lastigste. Wat ben ik blij dat jij er steeds voor mij was wanneer ik het even niet meer zag zitten om door te zetten. Bedankt, ik zie jou heel graag! **Marleen**, bedankt om zo goed voor mij te zorgen en lekker te koken in de vele weekends dat ik bij jullie in het bureau aan mijn doctoraat gewerkt heb. **Lisa**, bedankt om mij te vergezellen in het bureau en in mijn pauzes, vooral wanneer Bjorn niet thuis was. Samen met **Didier, Willem, Flore** en **Thomas**, bedankt voor de leuke familie momentjes en voor jullie motiverende woorden!

Tenslotte wil ik natuurlijk ook mijn familie bedanken. **Tantes** en **nonkels, neven** en **nichten** (en kindjes), bedankt voor de vele leuke familiefeestjes en motiverende babbeltjes! **Meme Agnes** en **meme Denise**, mijn twee super meme's, steeds paraat om een kaars te branden als geluksbrenger. Bedankt voor de vele leuke etentjes, koffienamiddagen, babbeltjes en telefoontjes! **Pepe Kamiel**, nu heb je al twee muziekconcerten van mij moeten missen en moet je ook dit missen. Bedankt voor de vele grapjes en het lekkere eten. Ik weet zeker dat je met veel trots toekijkt van daar boven. **Senna**, onze hond, bedankt voor het luide gehuil telkens als ik thuiskwam en de vele knuffeltjes wanneer nodig. **Silke**, mijn zus, bedankt voor jouw altijd overenthousiaste motiverende woorden en jouw steun (en knuffels) wanneer nodig. Bedankt ook voor onze fijne bezoeken aan de meme's samen en de avondjes samen koken en 'Thuis' kijken bij mij thuis. Zeker ook bedankt om dit dankwoord na te lezen en te controleren op spellingfouten! **Mama** en **papa**, bedankt voor de vele kansen die jullie mij doorheen mijn gehele leven reeds gegeven hebben en nog zullen geven. Bedankt ook voor jullie wijze raad waaruit ik al veel geleerd heb. Bedankt om mij ook steeds de juiste weg te wijzen wanneer ik deze even kwijt was. Bedankt voor de vele lekkere potjes eten of de vele keren dat ik bij jullie kwam eten, omdat ik weeral geen tijd had om te koken. Maar vooral bedankt voor jullie onvoorwaardelijke liefde en steun! Ik zie jullie, en de rest van de familie, graag!

BEDANKT! Lore (September 2017)

Curriculum vitae

Personal information

First name: Lore

Last name: Eggermont

Date of birth: November 7, 1988

Place of birth: Kortrijk

Nationality: Belgian

Marital status: not married

Address: Burgemeester Vercruysselaan 7/11, 8500 Kortrijk

Work address: Coupure Links 653, 9000 Ghent

Tel.: +32477442886

E-mail: lore.eggermont@gmail.com

Education

2014 – 2016: Qualification of teacher (CVO VIVO, Kortrijk)

2009 – 2011: Master of Science in Biochemistry and Biotechnology

Master thesis: The role of the polarisome in the thau pathology of a humanized yeast model

Promotor: Prof. Joris Winderickx (Functional Biology, Heverlee, KULeuven)

2006 – 2009: Bachelor of Science in Biochemistry and Biotechnology (KULeuven)

2000 – 2006: High School Degree – Science and Mathematics, Spes Nostra, Kuurne

Professional record

2011 – 2017: Assistant and PhD candidate at the Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University

PhD thesis: Nictaba homologs from *Arabidopsis thaliana* and their role in plant stress responses

Promotor: Prof. dr. Els J.M. Van Damme

Publications

Eggermont L, Verstraeten B, Van Damme EJM. 2017. Genome wide screening for lectin motifs in *Arabidopsis thaliana*. *The Plant Genome* 10, 1-17.

Eggermont, L., Stefanowicz K., Van Damme EJM. 2017. Nictaba homologs from *Arabidopsis thaliana* are involved in plant stress responses. *Frontiers in Plant Science*. In review.

Van Holle S, De Schutter K, **Eggermont L**, Tsaneva M, Dang L, Van Damme EJM. 2017.

Comparative study of lectin domains in model species: new insights into evolutionary dynamics. *International Journal of Molecular Sciences* 18, 1136.

Stefanowicz K, Lannoo N, Zhao Y, **Eggermont L**, Van Hove J, Al Atalah B, Van Damme EJM. 2016. Glycan-binding F-box protein from *Arabidopsis thaliana* protects plants from *Pseudomonas syringae* infection. *BMC Plant Biology* 16, 213.

Van Hove J, Stefanowicz K, De Schutter K, **Eggermont L**, Lannoo N, Al Atalah B, Van Damme EJM. 2014. Transcriptional profiling of the lectin ArathEULS3 from *Arabidopsis thaliana* toward abiotic stresses. *Journal of Plant Physiology* 171, 1763-1773.

Conferences

Conferences with oral presentation

Eggermont L. 2015. Elucidating the function of Nictaba homologues in *Arabidopsis thaliana*. Plant research FBW Meeting 2015, March 19, Ghent, Belgium.

Conferences with poster presentation

Van Damme EJM, Delporte A, **Eggermont L**, Lannoo N. 2012. The sweet side of histone proteins: interaction with lectins. Keystone Symposia: nuclear events in plant gene expression and signalling, Sagebrush Inn & Suites, March 6-11, Taos, New Mexico, USA.

Eggermont L, Delporte A, Van Damme EJM. 2012. Lectin-histone interactions in the plant cell nucleus. International Conference on Arabidopsis Research 2012, Hofburg Imperial Palace, July 3-7, Vienna, Austria.

Eggermont L, Stefanowicz K, Van Hove J, Lannoo N, Van Damme EJM. 2013. Expression analysis of Nictaba homologues in *Arabidopsis thaliana* subjected to abiotic and biotic stresses. 24th Joint Glycobiology Meeting, Leucorea, November 24-26, Wittenberg, Germany.

Eggermont L, Stefanowicz K, Van Damme EJM. 2014. Expression analysis of Nictaba homologues in *Arabidopsis thaliana* subjected to abiotic stresses. 25th Joint Glycobiology Meeting, Het Pand, September 14-16, Ghent, Belgium.

Eggermont L, Smagghe G, Van Damme EJM. 2016. Nictaba homologues in Arabidopsis: a plant defense army?! Tri-National Arabidopsis Meeting 2016, BOKU, September 14-16, Vienna, Austria.

Conferences without poster presentation

206th meeting of the Belgian Society for Biochemistry and Molecular Biology: Proteomics and Interactomics in Signalling Networks, 2012, May 25, Ghent, Belgium.

Plant research FBW Meeting 2014, February 5, Ghent, Belgium.

