Identification of Peroxisome-Targeted Proteins Implicated in Plant Innate Immunity in *Arabidopsis thaliana*

By

Amr Ramzy Abass Kataya

Thesis submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy
(PhD)

Faculty of Science and Technology
Department of Mathematics and Natural Science

2011

University of Stavanger
N-4036 Stavanger
Norway
www.uis.no

ISBN
ISSN

Table of contents

Acknowledgements

I would like to express my deep gratitude to my supervisor Prof. Sigrun Reumann, for the guidance and scientific support she has provided throughout the course of this study, and specially her valuable comments and constructive criticism.

Special thanks to University of Stavanger for giving me this honor to do my PhD studies and to be a member of its research endeavors. Moreover, very special thanks for Prof. Peter Ruoff, Prof. Cathrine Lillo, Prof. Bjørn Hjertager, and Prof. Simon G. Møller for their kindness and support. I would be grateful also to mention Mrs. Elisabeth Stornes Fiskå, TN faculty administration advisor for her much guidance and helpful information.

I would like also to express my gratefulness for Prof. Jianping Hu and Prof. Sheng Yang He from PRL/University of Michigan/USA for giving me the opportunity to do part of my research at their labs. Moreover, I would like to thank Dr. Matt Oney, Dr. Gaelle Cassin, and Dr. Francisco J. Uribe, and all the members from both groups located at MSU-PRL/USA. I wish also to thank Dr. Jodi Maple and Dr. Melinda Frame/MSU/USA for their valuable microscopy training and guidance. I wish also to thank Dr. Tanja Meyer/Münster/Germany for teaching me transient expression using protoplasts.

It would be a long list to mention all friends and colleagues I am indebted to. It is my great pleasure to thank all of them for their support. Special thanks to my collaborators and friends at CORE: Dr. Xiong-yan Chen, Pradeep Soni, Aline Benichou, Chimuka Mwaanga; Altinai Adilbayeva, Gopal Chowdhary, Behzad Heidari Ahootapeh, Dugassa Nemie-Feyissa, Dr. Kristine Marie Olsen, Dr. Else Muller Jonassen, Karène Jacques Jensen, Ingunn Jolma, Mohammed Gebriel, Prof. Lutz Eichacker, Dr. Xiang Ming Xu, Dr. Daniela Gargano, Dr. Benny Björkblom, Dominik Piston, Janine Arnold, and Xiao-Yu Ni.

Finally, I would like to express my deepest love and recognition to my father, mother, wife and my brothers, for their love and continuous support. I owe them everything I have today and I dedicate this thesis to them. Furthermore, I would like to express my deepest impatience for waiting to see my first child, whom shall come to this life "enshaa Allah" with the beginning of November/2011, and I dedicate this thesis to him or her.

Abstract

Peroxisomes are subcellular organelles, traditionally known to be involved in processes like photorespiration, fatty acid β-oxidation, and detoxification of reactive oxygen species. Proteome analysis of plant peroxisomes and targeting signal prediction methods are important tools to identify novel peroxisomal proteins. In the present study the accuracy of newly developed methods to predict peroxisome targeting signals type 1 (PTS1) in plant proteins was investigated by *in vivo* subcellular targeting analyses. Upon application of these prediction methods to the *Arabidopsis thaliana* genome, 392 gene models were predicted to possess functional PTS1 domains, several proteins of which were validated as peroxisomal and numerous novel PTS1 tripeptides were identified. Furthermore, several detoxification-related enzymes and defense-related *Arabidopsis* proteins were detected by proteome analyses and PTS1 prediction methods that were potentially targeted to peroxisomes.

Two enzymes of the ascorbate-glutathione (ASC-GSH) cycle, glutathione reductase 1 (GR1) and dehydroascorbate reductase 1 (DHAR1), and five glutathione-S transferases (GSTs) had been detected by proteome analysis in leaf peroxisomes. *In vivo* subcellular localization targeting analyses of the present study verified peroxisomal targeting for GR1 and the protein was found to carry a functional novel PTS1 (TNL>). By contrast, the four GSTs remained cytosolic in the chosen orientation in the back of the reporter protein.

New fragmented evidence has been emerging in the literature for an important role of plant peroxisomes in innate immunity. In the present study sixteen defense-related *Arabidopsis* proteins were experimentally investigated for protein targeting to peroxisomes by *in vivo* subcellular localization. The proteins of interest included several yet unknown homologs of *Arabidopsis* NDR1 and tobacco HIN1, the so-called NDR1/HIN1 like (NHL) proteins. *In vivo* subcellular localization was primarily investigated for three NHL family members (NHL4, NHL6 and NHL25). Peroxisome targeting was verified for NHL4 with strong indications also for NHL6 and NHL25 in being located in peroxisomes. AtIAN12 is a homolog of AIG1/AtIAN8 and had been identified by *Arabidopsis* leaf peroxisome proteomics. *In vivo* subcellular localization experiments demonstrated that AtIAN12 protein is targeted

to peroxisomes and indicated that the targeting pathway involves post-translational protein modification by isoprenylation. Taken together, the data indicate for the first time that one NDR1/HIN1 homolog (NHL4) and AtIAN homolog (AtIAN12) are peroxisome associated.

Preliminary gene expression analyses indicated that three *NHL* genes and three *AtIAN* genes are induced by a bacterial pathogen (*Pst* DC3000), while *NHL6*, *NHL25*, and *AtIAN8* are induced by an avirulent *Pst* DC3000 strain (carrying the effector avrRpt2). Out of the six *NHL* and *AtIAN* genes, only *NHL6* appeared to be induced in wt Col-0 plants by the bacterial elicitor (flg22), but remained unaffected in *Arabidopsis* plants carrying a mutation in the flagellin receptor gene *FLS2*. The data suggested that NHL6 is involved in basal PAMP triggered immunity (PTI). Furthermore, *NHL6* transcripts accumulated similarily in both wt plants and *npr1* mutant plants after flg22 treatment, which indicates that *NHL6* induction is NPR1-independent. Functional studies were initiated through the isolation of homozygous mutants, amiRNA lines and overexpresser lines for selected *NHL* and *AtIAN* genes. In homozygous mutants (three *nhl* mutants and *ian11*), differences in bacterial proliferation were observed compared to wt plants upon infection with the avirulent bacterium *Pst* DC3000 (avrRpt2). Overall, the identification of several defense-related proteins in peroxisomes together with preliminary functional data on NHL proteins opens new perspectives to important, multi-layered peroxisome functions in plant innate immunity.

List of publications

1- LINGNER, T., KATAYA, A.R., ANTONICELLI, G.E., BENICHOU, A., NILSSEN, K., CHEN, X.Y., SIEMSEN, T., MORGENSTERN, B., MEINICKE, P. & REUMANN, S. (2011) Identification of Novel Plant Peroxisomal Targeting Signals by a Combination of Machine Learning Methods and *in Vivo* Subcellular Targeting Analyses. *Plant Cell*, **23,** 1556-1572.

2- KATAYA, A.R.A. & REUMANN, S. (2010) Arabidopsis Glutathione Reductase 1 Is Dually Targeted to Peroxisomes and the Cytosol. *Plant Signal Behav*, **5:2,** 171-175.

List of abbreviations

AIG	avrRpt2-induced gene
APS	Ammonium persulfate
APX	Ascorbate peroxidase
ASC-GSH	Ascorbate-glutathione
AtSurE	*Arabidopsis* acid phosphatase survival protein SurE
avr	Avirulence
CAD7	Cinnamyl-alchol dehydrogenase7
CaMV	*Cauliflower mosaic virus*
CaMV 35S promoter	*Cauliflower mosaic virus* 35s promoter
CC	Coiled coil
CDR1	Constitutive disease resistance 1
CFP	Cyan fluorescence protein
CFU	Colony forming units
CoA	Coenzyme A
coxIV	Cytochrome c oxidase IV subunit
DAPI	4', 6-diamidino-2-phenylindole
DHAR	Dehydroascorbate reductase
DMSO	Dimethyl sulfoxide
DRP	Disease resistance protein-related
EDS1	Enhanced disease susceptibility 1
EDTA	Ethylene diaminetetra acetic acid
EFR	Elongation factor-Tu receptor
ER	Endoplasmic reticulum
EST	Expressed sequence tag
ETI	Effector-triggered immunity
EYFP	Enhanced yellow fluorescent protein
FLS2	Flagellin sensing 2
GFP	Green fluorescent protein
GIMAP	GTPase of immunity-associated proteins
gMDH	Glyoxysomal malate dehydrogenase
GR	Glutathione reductase
GST	Glutathione S-transferase
GTPases	GTP-binding proteins
H_2O_2	Hydrogen peroxide
HIN1	Harpin-induced gene1
His	Histidine
HR	Hypersensitive response
IAA	Indole acetic acid
IAN	Immune-associated nucleotide-binding proteins
IBA	Indole-3-butyric acid
IPTG	Isopropyl-beta-D-thiogalactopyranoside
JA	Jasmonic acid
LB	Lysogeny broth

LIMDP	LIM domain-containing protein
LM	Low salt medium
LPS	Lipopolysaccharides
LRR	Leucine rich repeat
LS	Linsmaier & Skoog
MAPK	Mitogen-activated protein kinase
MBP	Maltose binding protein
MDAR	Monodehydroascorbate reductase
MIF	Macrophage migration inhibitor factor
MS	Murashige and Skoog
NBS	Nucleotide binding site
NDR1	Non-race specific disease resistance 1
NHL	NDR1/HIN1 like proteins
NPR1	Nonexpressor of PR genes 1
OFP	Orange fluorescent protein
OZI1	Ozone-induced protein 1
PAMP	Pathogen-associated molecular pattern
PEG	Polyethylene glycol
PEN	Penetration
PEX	Peroxin
PM	Plasma membrane
PMP	Peroxisomal membrane protein
PMSF	Phenylmethylsulfonyl fluoride
PPT	Phosphinotricin
PR	Pathogenesis-related
PRR	Pattern recognition receptor
Ps	*Pseudomonas syringae*
Pst DC3000	*Pseudomonas syringae* pv. tomato DC3000
PTD	Peroxisomal targeting domain
PTI	PAMP-triggered immunity
PTS	Peroxisome targeting signal
pv.	Pathover
R	Resistance
RFP	Red fluorescent protein
RIN4	RPM1-interacting protein 4
ROS	Reactive oxygen species
RPM	Resistance to *Pseudomonas syringae* pv. maculicola
RPS	Resistance to *Pseudomonas syringae*
RuBisCO	Ribulose bisphosphate carboxylase/oxygenase
SA	Salicylic acid
SAR	Systemic acquired resistance
TEMED	N'-tetramethylethane-1,2-diamine
TIR	Toll interleukin receptor
TTSS	Type III secretion system
X-Gal	5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside

List of figures

List of tables

1. Introduction

1.1 Peroxisomes

Peroxisomes are single membrane bound subcellular organelles, present in all major groups of eukaryotes (Gabaldon, 2010). They are usually spherical microbodies in the range of 0.1 to 1 µm in diameter. Peroxisomes were first discovered as compartments containing hydrogen peroxide (H_2O_2) generating oxidases together with catalase that degrades H_2O_2 into molecular oxygen and water (De Duve and Baudhuin, 1966; van den Bosch et al., 1992; Kaur et al., 2009). Fatty acid β-oxidation and H_2O_2 detoxification are two well conserved functions of peroxisomes, but specialized functions were also identified, for example plant glyoxysomes are specialized peroxisomes in germinating seeds that harboring the glyoxylate cycle (Escher and Widmer, 1997; Graham, 2008). Plant leaf peroxisomes take part in photorespiratory glycolate metabolism, and the biosynthesis of hormones [indole acetic acid (IAA), Salicylic acid (SA), and jasmonic acid (JA)] (Nyathi and Baker, 2006). Glycosomes are found in trypanosomes where they contain the enzymes for glycolysis (Michels, 1988). Yeast peroxisomes are equipped with enzymes for methanol and amine oxidation (Veenhuis et al., 1983). Mammalian peroxisomes carry the enzymes involved in lipid and cholesterol synthesis (Wanders and Waterham, 2006; Wierzbicki, 2007).

1.1.1 Plant peroxisome functions

1.1.1.1 Metabolic functions

In peroxisomes, the β-oxidation pathway is responsible for fatty acid degradation. First, fatty acids are imported into peroxisomes and activated to coenzyme A (CoA) esters that are processed by sequential cleavage through β-oxidation. There are several physiological roles of β-oxidation in plants, for example embryo and flower development and production of signaling molecules [JA, SA and IAA, (Poirier et al., 2006; Kaur et al., 2009)].

During germination, long-chain fatty acids (which form triacylglycerol reserves in oil bodies in *Arabidopsis* and other oilseed plants) are used to provide energy. The degradation of triacylglycerols is carried out by glyoxysomes. Initially, triacylglycerols are activated by CoA and further degraded in successive steps and converted to acetyl-CoA. Acetyl-CoA is then converted to succinate, which is transported to the mitochondria where it fuels the tricarboxylic acid cycle (Figure 1.2) and to produce sucrose (Gerhardt, 1992; Olsen, 1998).

Almost five decades back, photorespiration was discovered in isolated spinach leaf peroxisomes after the detection of glycolate oxidase through the production of glycine from [^{14}C] glycolate, as glycolate is converted to glyoxylate with the production of H_2O_2. Glyoxylate is then transaminated and converted to glycine (Kisaki and Tolbert, 1969; Tolbert et al., 1969). Photorespiration is initiated by the oxygenase reaction of ribulose bisphosphate carboxylase/oxygenase (RubisCO) which is functioning according to O_2 concentration and light intensity. Photorespiration is coordinated across chloroplasts, peroxisomes, and mitochondria. Briefly, photorespiration is initiated when RuBisCO is activated in chloroplasts and produces two moles of phosphoglycolate (byproducts of the oxygenase reaction) which are converted to one mole of phosphoglycerate (intermediate of the Calvin–Benson cycle), and one CO_2 by the photorespiratory glycolate pathway (Hayashi and Nishimura, 2006).

Leaf peroxisomes convert glycolate to glycine and serine to glycerate by the enzymes glycolate oxidase, hydroxypyruvate reductase and two aminotransferases. The glycine produced is subsequently converted to serine in mitochondria by decarboxylation by glycine decarboxylase and serine hydroxymethyl transferase with the production of ammonia as a byproduct. Serine then re-enters the peroxisome to be transaminated by serine-glyoxylate aminotransferase to yield hydroxypyruvate, which is reduced by NADH (provided by peroxisomal malate dehydrogenase) to glycerate in a reaction catalyzed by hydroxypyruvate reductase. Finally, glycerate is phosphorylated in the chloroplast by a stromal glycerate kinase to produce 3-phosphoglycerate, which feeds into the Calvin cycle [Figure 1.1, (Hayashi and Nishimura, 2006; Reumann and Weber, 2006; Kaur et al., 2009)].

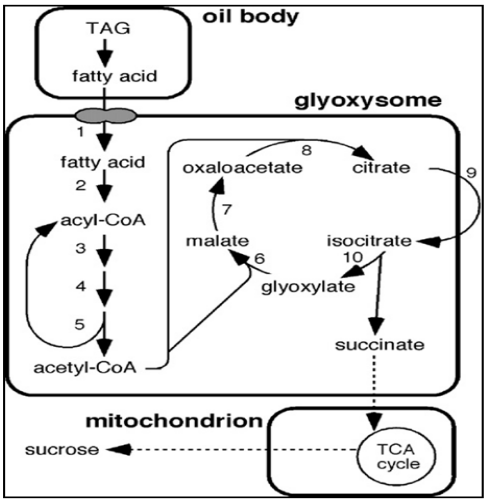


Figure 1.2: Gluconeogenesis from seed fatty acids
The conversion of fatty acids to succinate takes place in glyoxysomes via fatty acid β-oxidation (1–5) and the glyoxylate cycle (6–10). The enzymes involved in these pathways are: 1, full size ABC transporter; 2, acyl-CoA synthetase; 3, long-, medium- and short chain acyl-CoA oxidases; 4, the multifunctional protein possessing enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase activities; 5, 3-ketoacyl-CoA thiolase; 6, malate synthase; 7, malate dehydrogenase; 8, citrate synthase; 9, aconitase; 10, isocitrate lyase. Figure taken from (Hayashi and Nishimura, 2006).

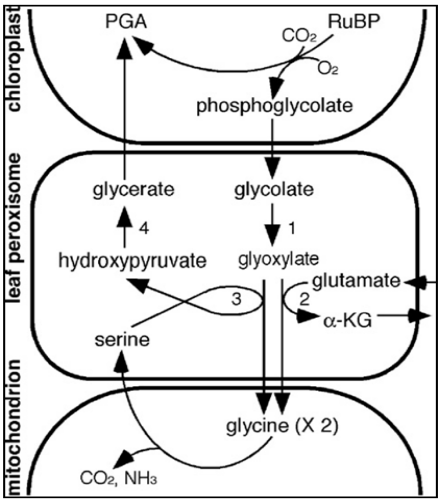


Figure 1.1: Photorespiratory glycolate metabolism
Photorespiration in photosynthetic tissue of C3 plants. Within the entire photorespiratory glycolate pathway, the leaf peroxisome converts glycolate to glycine and serine to glycerate. The enzymes involved in this metabolism are: 1, glycolate oxidase; 2, glutamate-glyoxylate aminotransferase; 3, serine-glyoxylate aminotransferase; 4, hydroxypyruvate reductase. Figure taken from (Hayashi and Nishimura, 2006).

1.1.1.2 Detoxification functions

Peroxisomes are also involved in the production of ROS [e.g. H_2O_2 and superoxide radicals (O2•−)], and reactive nitrogen species. These molecules are implicated in intra- and inter-cellular signaling. Under normal conditions equilibrium exists between the rate of synthesis and degradation of these molecules, while different biotic and abiotic stresses are disturbing this balance, which may initiate a signaling cascade or cause cellular damage, see 1.1.1.3, (Corpas et al., 2001; Nyathi and Baker, 2006).

The H_2O_2 produced in peroxisomes is degraded by antioxidant enzymes: catalase and ascorbate-glutathione (ASC-GSH) cycle

enzymes. The inactivation of peroxisomal antioxidant enzymes could create toxic conditions in the plant cell, leading to oxidative damage and cell death. For example catalase is reported to be inactivated by high light, peroxynitrite and different stress conditions (Corpas et al., 2001; Reumann and Corpas, 2010). When catalase is inactivated in peroxisomes, the ASC-GSH cycle is another alternative for H_2O_2 degradation. The ASC-GSH cycle had been described only biochemically in pea peroxisomes (Jimenez et al., 1997). While ascorbate peroxidase (APX) 3 and monodehydroascorbate reductase (MDAR) 1 and 4 had been cloned and validated to be peroxisomal proteins (Leterrier et al., 2005; Lisenbee et al., 2005; Narendra et al., 2006). Glutathione reductase (GR) and dehydroascorbate reductase (DHAR) had not been cloned from any plant species but only been biochemically characterized and found in peroxisomal proteome studies (Jimenez et al., 1997; Reumann et al., 2007; Reumann et al., 2009). According to the proposed model of ASC-GSH cycle, the membrane-bound APX in collaboration with MDAR, degrades H_2O_2 that can diffuse out of peroxisomes, as well as H_2O_2 being formed by $O2•−$ dismutation at the cytosolic side of the peroxisomal membrane. DHAR and GR, located in the peroxisomal matrix, accomplish detoxification of H_2O_2 produced in the matrix in sequential ascorbate- and glutathione-dependent reactions. The ASC-GSH cycle also provides NAD^+ for peroxisomal metabolism and GSH protects the flavin-containing oxidases against photo-inactivation (Jimenez et al., 1997; Reumann and Corpas, 2010).

Other important key factors in detoxification reactions are glutathione S-transferases [GSTs; 48 members classified into phi, tau, theta, zeta and lambda classes, (Edwards and Dixon, 2005)]. GSTs have several glutathione (GSH) dependent functions including the conjugation and resulting detoxification of herbicides, the reduction of organic hydroperoxides formed during oxidative stress and others, e.g. soluble GSTs act as glutathione peroxidases (Edwards and Dixon, 2005). Three members of the GST subfamily theta (T) have been shown to be peroxisome-targeted (Reumann et al., 2007; Dixon and Edwards, 2009). Additionally, four GSTs of the U and F subfamilies (GSTU19, GSTU20, GSTF7 and GSTF10) have been identified in isolated *Arabidopsis* leaf peroxisomes by proteome analyses (Reumann et al., 2009).

1.1.1.3 Stress-related functions

Essential roles of peroxisomes in stress and plant defense responses against pathogens were recently reported. For example, peroxisomes proliferate much more under different stress conditions produced by xenobiotics, ozone, heavy metals, wounding, salt and pathogen attack (Mitsuya et al.; Corpas et al., 2001). Additionally, two *Arabidopsis* small heat-shock proteins were identified in peroxisomes and one of them was reported to be induced by heat and oxidative stress, which supports the proposed roles of peroxisomes in stress responses (Ma et al., 2006; Kaur et al., 2009). Peroxisome biogenesis genes (PEX, see 1.1.2) were also reported to be induced by physiological elevated H_2O_2 which is produced during stress conditions in response to wounding and to infection with avirulent bacteria (Lopez-Huertas et al., 2000). Moreover, by monitoring antioxidant enzyme activities in isolated leaf peroxisomes that were isolated from tomato leaf cells (infected by the necrotrophic fungus, *Botrytis cinerea*), the peroxisomal antioxidant system as a hole was found to be significantly affected. During early stages, the activities of peroxisomal enzymes such as superoxide dismutase, glutathione peroxidase and catalase increased, while they decreased at later stages (Kuzniak and Sklodowska, 2005). In the same study, the peroxisomal ASC-GSH cycle enzyme activities were reported to be decreased by infection without any activity increase at earlier stages. These data indicate that the collapse of the antioxidant system might be important for pathogen-induced cell death (Kuzniak and Sklodowska, 2004, 2005).

With last decade discoveries, the function of peroxisomes in plant defense responses against pathogens (see 1.2) started to be reported. For example, one wild melon line gained resistance to an oomycete pathogen (*Pseudoperonospora cubensis*; causing foliar disease of cucurbit) due to the overexpression of a peroxisomal photorespiratory aminotransferase (Taler et al., 2004). Moreover, peroxisomes were detected to migrate and accumulate at pathogen infection sites in two different studies (Koh et al., 2005; Lipka et al., 2005), which might provide a mechanism for the activation and release of toxic molecules at a high concentration. During powdery mildew (*Erysiphe cichoracearum*) infection of *Arabidopsis* epidermal cells, organelles including peroxisomes moved towards and accumulated at fungi penetration sites, Figure 1.3, A, (Koh et al., 2005). During research studies on penetration 2 (PEN2) protein, PEN2-labeled peroxisomes

were reported to accumulate at fungal (*Blumeria graminis* f. sp. Hordei) entry sites (Figure 1.3, B). Furthermore, mechanical wounding by fine needle penetration led to the accumulation of peroxisomes at the penetration site. These data indicate that mechanical wounding simulates pathogen penetration and induces preinvasion defense mechanism (Figure 1.3, C1-3) (Hardham et al., 2008).

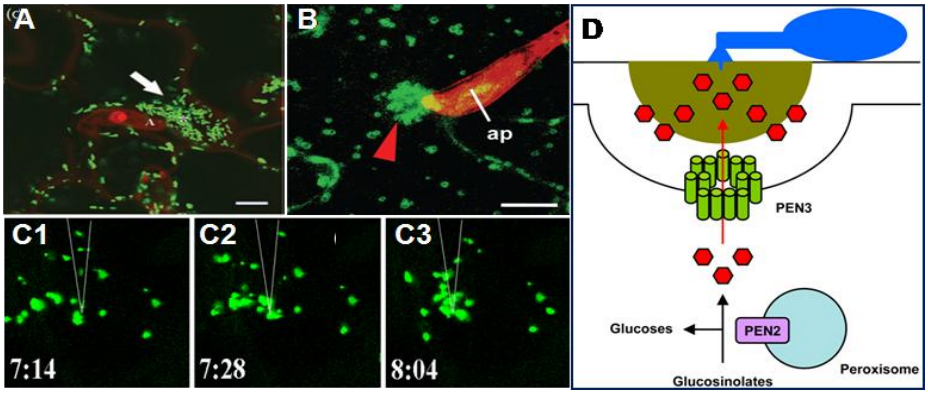


Figure 1.3: Peroxisome role in plant innate immunity
A: Aggregation of GFP-labeled peroxisomes at *Erysiphe cichoracearum* infection sites: image taken from (Koh et al., 2005). B: PEN2-GFP-labeled peroxisomes at condidiospore entry site: image taken from (Lipka et al., 2005). C: GFP-labeled peroxisomes at micro-needle penetration site: images taken from (Hardham et al., 2008). D: Pathogen-triggered and ABC transporter-driven efflux of small molecules into the apoplast in response to infection. In *Arabidopsis*, PEN3 is required for pre-invasive resistance to a broad range of fungal parasites: images taken from (Kwon et al., 2008).

PEN2 is a peroxisomal glycosyl hydrolase that is essential in inducible pre-invasion resistance mechanism, and its loss in *pen2* plants led to increased susceptibility to *Blumeria graminis* f. sp. Hordei (Lipka et al., 2005). PEN2 possesses myrosinase activity to initiate pathogen-triggered metabolism of indole glucosinolates cleaving glucose from thioglucosides (Grubb and Abel, 2006; Bednarek et al., 2009). Moreover, PEN2 and pathogenesis-related (PR) proteins were induced together with callose after the treatment by bacterial derived elicitors e.g. flg22, see 1.2.1.1 (Gomez-Gomez et al., 1999). PEN2 was also found to be a crucial component for callose deposition, as *pen2* mutants failed to display flg22-induced callose deposition (Clay et al., 2009; Kaur et al., 2009). Callose is an amorphous, high-molecular-weight β-1,3-glucan and is deposited in cell wall appositions (papillae) that form beneath infection sites. Callose is thought to provide a physical barrier

to pathogen penetration (Gomez-Gomez et al., 1999; Luna et al., 2011). These data suggest that PEN2 might function as a signalling molecule or co-activator in flg22-induced callose deposition (Clay et al., 2009; Kaur et al., 2009). In contrary, callose deposition increased dramatically (Luna et al., 2011) in the *cat2-1* mutant which accumulates high levels of H_2O_2 (Bueso et al., 2007). The role of elevated H_2O_2 production in peroxisomes is suggested to be responsible for conferring resistance against pathogen infection (Heath, 2000; Taler et al., 2004; Kaur et al., 2009).

It was studied that PEN3 (an ATP-binding cassette-type (ABC) transporter), similarly to PEN1, accumulates and associates beneath plasma membrane (PM) sites infected by a fungus (Figure 1.3, D). PEN1 is a syntaxin that belongs to the superfamily of soluble N-ethylmaleimide sensitive factor attachment protein receptor (SNARE) domain-containing proteins (Assaad et al., 2004; Stein et al., 2006). PM-associated PEN3 was suggested to translocate PEN2-generated molecules into the apoplastic space (Figure 1.3, D). These proteins (PEN 2 and 3), most likely together, constitute a dedicated secretory immune response pathway for small molecules with broad-spectrum antimicrobial activity (Lipka et al., 2005; Kwon et al., 2008; Bednarek et al., 2009; Bednarek and Schulze-Lefert, 2009; Bednarek et al., 2010).

1.1.2 Peroxisome biogenesis

Peroxisomal proteins are nuclear-encoded, synthesized on cytosolic ribosomes, and the proteins are transported into peroxisomes with the help of peroxins, PEX (encoded by *PEX* genes). Peroxins function in different processes of peroxisome biogenesis such as peroxisome *de novo* biogenesis, import of proteins and peroxisome proliferation (Orth et al., 2007). For a long time, peroxisomes were viewed as semiautonomous organelles that exist outside the secretory and endocytic pathways of vesicular flow. Recently, it has become clear that peroxisomes are derived from the endoplasmic reticulum (ER) although they also multiply by proliferation (Hoepfner et al., 2005). Two groups of peroxisomal membrane proteins (PMPs, Figure 1.4) were suggested (I and II) based on their import pathways. Group I PMPs are inserted post-translationally into the ER membrane after being synthesized in the cytosol, and then transported to peroxisomes via specific ER vesicles. Group II PMPs are sorted to peroxisomes

directly from the cytosol. Peroxisome membrane import depends on membrane PTSs (mPTSs) that have been identified in group I and II PMPs (Hoepfner et al., 2005; Kragt et al., 2005; Mullen and Trelease, 2006). Peroxisome proliferation by division was also reported and divided into three steps including elongation (by PEX11), membrane constriction and final fission steps by fission and dynamin-related proteins (Orth et al., 2007; Lingard et al., 2008; Kaur et al., 2009).

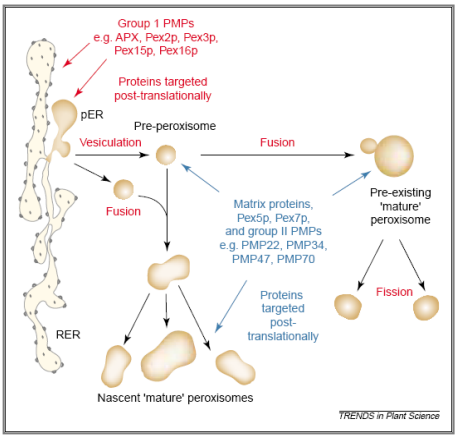


Figure 1.4: Proposed role of the ER in peroxisome biogenesis
Group I PMPs, including APX and various peroxins, are sorted to the rough ER (RER) and/or peroxisomal ER (pER). Pre-peroxisomes might sort to and fuse with a pre-existing mature peroxisome, or they might fuse with other pre-peroxisomes to form nascent mature peroxisomes. All the post-ER peroxisomal compartments are capable of post-translational uptake of matrix proteins and certain (group II) PMPs: Figure taken from (Mullen et al., 2001).

1.1.3 Matrix protein import into peroxisomes

Peroxisomal matrix proteins are encoded in the nucleus and translated in the cytosol before being transported to peroxisomes. Matrix proteins are targeted to their destination by a peroxisome targeting signal (PTS). A major breakthrough in the elucidation of the mechanism of protein import into peroxisomes was the identification of the PTS type 1 (PTS1; conserved tripeptide, SKL>) at the C-terminus of luciferase of the firefly *Photinus pyralis* (Gould et al., 1987; Gould et al., 1989). The majority of the identified peroxisomal matrix proteins has a PTS1, while some proteins have an N-terminal PTS type 2 (PTS2) which is a nonapeptide with RLx$_5$HL as the prototype sequence (Kaur et al., 2009). The PTS1- or PTS2-containing matrix proteins are recognized by soluble receptors, PTS1 by PEX5 (Figure 1.5, A), and PTS2 by PEX7 (Figure 1.5, B) in the cytosol, that guide them to a docking site at the peroxisomal membrane (Kaur et al., 2009; Lingard et al., 2009). *Arabidopsis* PEX5 and PEX7 interact with each other, and silencing experiments of PEX5 and PEX7 transcripts show that PEX7 is required for PTS2 protein import, whereas reducing PEX5 affects both PTS1

and PTS2 protein import (Nito et al., 2002; Baker and Sparkes, 2005; Khan and Zolman, 2010).

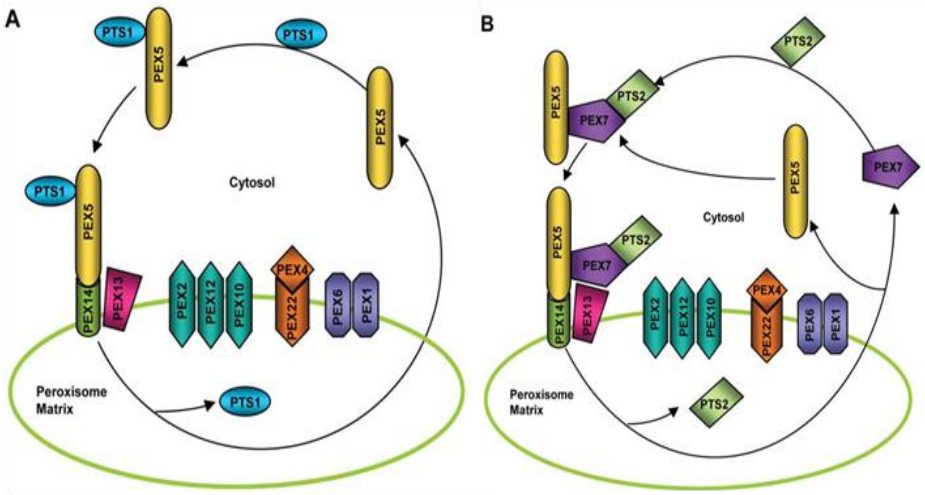


Figure 1.5: Representative model for matrix protein import
(A) PTS1 protein import. PEX5 recognizes and binds PTS1-containing proteins in the cytosol. (B) PTS2 import. PEX7 recognizes and binds PTS2-containing proteins in the cytosol: Figures taken from (Kaur et al., 2009).

1.1.4 Tools for identification of the peroxisome proteome

The identification of new PTS1 and PTS2 peptides of peroxisomal proteins was enlarged in the past two decades. Firstly, PTS1 sequence characteristics were experimentally revealed as a small uncharged-basic-nonpolar>, [SAC][KRH]L> (Gould et al., 1989; Swinkels et al., 1992). PTS1-specific variations were studied by comparing the peroxisome targeting efficiency of β-glucuronidase constructs which had several different C-terminus tripeptides and identified the first plant-specific PTS1 consensus sequence (referred here to as the Hayashi motif; [CASP][KR][ILM]>) (Hayashi et al., 1996; Hayashi et al., 1997). In another study, a more permissive consensus motif ([ACGST][HKLNR][ILMY]>) was reported, during experimental verifications by the reporter protein chloramphenicol acetyltransferase and transient expression in BY-2 suspension-cultured cells of *Nicotiana tabacum* (Mullen et al., 1997). Moreover, based on the interaction of PTS1 tripeptides with tobacco PEX5 in yeast two-hybrid system, a proline residue was identified in the pos. -3 and the motif became even

more permissive ([ACGPST][HKLNR][ILMY]>, (Kragler et al., 1998)). Recently, by combination of peroxisomal protein identification by proteome analysis (see 1.1.4.2) and *in vivo* subcellular localization studies, several PTS1s were established (SSL>, SSI>, ASL>, SLM>, and SKV>) which was inferring a new non-basic residue (S) at pos. -2 (Reumann et al., 2007; Reumann et al., 2009). Additionally, many novel *Arabidopsis* PTS1 containing proteins were identified (Ma et al., 2006; Reumann et al., 2007; Eubel et al., 2008; Moschou et al., 2008; Kaur et al., 2009; Reumann et al., 2009; Babujee et al., 2010; Quan et al., 2010).

Many physiological functions of plant peroxisomes are difficult to study because of their fragile nature to handle *in vitro*. Moreover, as explained in section 1.1.1.3, peroxisomes are reported to have many low-abundance and stress-related proteins that are targeted to peroxisomes under special conditions. To identify new functions of plant peroxisomes, the determination of the peroxisome proteome is crucial. Three major methodologies have been applied to such studies: 1) bioinformatics-based prediction of PTS, 2) experimental peroxisome proteome analyses, and 3) experimental verification of putative peroxisomal proteins by *in vivo* subcellular localization. Indeed, *in vivo* subcellular localization studies were mostly applied to validate putative peroxisomal proteins that were detected from methodologies 1 and 2. The complete genome sequence of *Arabidopsis* (Arabidopsis genome initiative, 2000) facilitated screening of conserved PTS, and characterization of the peroxisomal identified proteins from experimental peroxisome proteome (Kaur et al., 2009).

1.1.4.1 Prediction of targeting signals

Bioinformatics-based predictions generally use mathematical models to predict targeting signals from genome sequences. Bioinformatics approaches were largely improved in identification of peroxisomal proteins based on known PTS1s (see 1.1.4). However, the predictions are facing some challenges (Kaur et al., 2009), for example the PTS1 and PTS2 might be undetectable, presence of alternative targeting signals, or targeting by "piggy-backing" on other proteins bearing PTSs (Purdue and Lazarow, 2001). Previous attempts to predict peroxisomal localization include PSORT, a knowledge-based predictor using a decision tree to sort proteins among several different compartments. In PSORT, the PTS1 motif [AS]-[HKR]-L is used as a marker for

peroxisomal location along with amino acid composition over the entire protein (Nakai and Kanehisa, 1992). Support vector machine (SVM) was also applied to predict protein localization to several organelles including peroxisomes based on amino acid composition and sequence (Cai et al., 2002). A pattern-based method including PTS1 and PTS2 motifs was also used to scan *Saccharomyces cerevisiae* ORFs for peroxisomal proteins. The authors were able to identify 18 putative peroxisomal proteins, where 10 of them were validated by subcellular localization studies (Geraghty et al., 1999). Another way to predict PTS1 proteins is to use the PROSITE pattern [ACGNST]-[HKR]-[AFILMVY] (Falquet et al., 2002). Other attempts were applied by combining prediction of PTS1s with domain-based cross-species comparisons. This combination significantly inferred higher specificity; PEROXIP [www.bioinfo.se/PeroxiP, (Emanuelsson et al., 2003)]. Other PTS1 predictors are also in use: PTS1 PREDICTOR [mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp, (Neuberger et al., 2003)], and PProwler [pprowler.itee.uq.edu.au, (Hawkins et al., 2007)]. Although several predictions are now available, plant-specific predictions still need much improvement because of the small and non-representative datasets (Reumann, 2004; Kaur et al., 2009).

Reumann (2004) assembled a true positive examples training dataset from PTS1-containing proteins. The assembled dataset was subjected to homology-based searches for the orthologs of peroxisomal proteins from the public protein sequence and expressed sequence tag (EST) databases. A 5-fold extension of the dataset of plant PTS1 proteins was gained after the usage of EST databases on *Arabidopsis* and identified novel non-canonical PTS1 tripeptides. According to this study, nine PTS1 tripeptides ([SA][RK][LM] without AKM> plus SRI> and PRL>) were identified in at least 10 sequences and three different groups were defined as major PTS1s. Moreover, eleven PTS1 tripeptides, including some unknown plant PTS1 tripeptides, were defined as minor PTS1s. A plant PTS1 tripeptide is predicted to be functional if it carries at least two of the six most abundant position-specific amino acid residues (i.e., S, A, R, K, L, M) in the form of [SA][RK]x>, [SA]y[LM]>, or z[RK][LM]> (Kaur et al., 2009).

Furthermore, seven to nine amino acid residues upstream of the tripeptide are important in enhancing or reducing the efficiency of targeting. Therefore, PTS1 protein prediction depends on both the probability for the C-terminal tripeptide to represent a functional PTS1

and the degree at which the upstream region (pos. -4 to -10 or -12) matches consensus PTS1 domains (Kaur et al., 2009). On average, minor PTS1 domains are enriched in basic residues, and/or hydrophobic residues [e.g., A, L, V, I, (Reumann, 2004; Kaur et al., 2009)]. Finally, PTS2 nonapeptides with RLx_5HL as the prototype were identified as restrictive PTS2s such as $R[ILQx_5HL]$ (Kato et al., 1996; Kato et al., 1998) or permissive PTS2s such as $[RK]x_6[HQ][ALF]$ (Flynn et al., 1998). Twelve functional PTS2s were characterized from the plant-specific EST training dataset of PTS2 proteins (Reumann, 2004).

1.1.4.2 Peroxisome proteomics

Experimental peroxisome proteome analyses were largely developed recently, after the improvement of peroxisome isolation methods (Kaur et al., 2009). Computational approaches helped in analyzing the data generated by mass spectrometry (ms) experiments, and to make predictions regarding the potential nature of the proteome. Several plant peroxisome studies were accomplished and helped to identify novel proteins from *Arabidopsis* (Kaur et al., 2009). Two proteome studies from *Arabidopsis* greening and etiolated cotyledons identified several known enzymes involved in ROS metabolism, photorespiration and fatty acid β-oxidation, where 33 out of 47 identified proteins from both studies were described as putative proteins of peroxisomes (Fukao et al., 2002; Fukao et al., 2003). In another two studies, new isolation protocols to purify leaf peroxisomes from *Arabidopsis* were developed, from which proteins were separated either by 1- or 2D gel electrophoresis. Peroxisomes were isolated from mature *Arabidopsis* leaves by Percoll density gradient followed by sucrose density gradient centrifugation followed by ms, and 42 out of 78 identified proteins were considered to be putative peroxisomal proteins. Seventeen proteins carried PTSs and eleven of them were validated as peroxisomal proteins by *in vivo* subcellular localization studies (Reumann et al., 2007). Peroxisomal protein identification was even doubled (150 proteins) after peroxisome enrichment through post-preparative immunoblotting analysis and by application of a 1DE shotgun ms approach. Fifty-five proteins were considered to be novel and 19 of them were validated by subcellular localization studies. SLM>, SKV> and RVx_5HF were also established as a new functional PTSs (Reumann et al., 2009). Other methods were applied to *Arabidopsis* suspension-cultured cells. Peroxisomes were purified by

free-flow electrophoresis and peroxisomal proteins were identified by two methodologies afterwards: (i) differential in-gel electrophoresis (DIGE) of enriched peroxisomes and mitochondria, and (ii) normalized spectral count analysis of shotgun proteome data from peroxisome fractions differing in their degree of purity. The identification of membrane proteins was optimized by sodium carbonate treatment of peroxisomes. Twenty of 89 identified proteins were considered to be novel (Eubel et al., 2008).

1.2 Plant defense responses

Plants are constantly subjected to attack by large numbers of bacteria, fungi, oomycetes, viruses and nematodes. It has been estimated that ~14% of crops produced worldwide are lost by plant diseases, accounting for more than £100 billion worldwide (Agrios, 2005). For example, late blight of potato caused by the oomycete *Phytophthora infestans*, resulted in a devastating epidemic in Northern Europe in the 1840s (Agrios, 2005). In addition, pathogen infections can affect negatively the quality of the crops by producing sometimes toxic compounds. For instance, the fungus *Claviceps purpurea* causes disease in cereals and grasses by producing toxic secondary metabolites in seeds that can be harmful for the consumer [e.g., ergot alkaloids, which are leading to ergotism in humans and animals (Keller et al., 2005)]. Recently, most of the research in plant disease resistance aims at finding broad-spectrum protection against infections. For this, understanding plant-pathogen interactions and the plant's defense mechanisms (which are referred to as plant innate immunity, see 1.2.1), might allow to improve or achieve engineered plant protection. Plant innate immunity is defined as the ability to recognize and respond to pathogens, and provides immediate defense against infection (Jones and Dangl, 2006).

Plant pathogens enter the apoplast (intercellular space) through different means. In general, bacteria use wounds or natural openings such as stomata whereas fungi and oomycetes simply penetrate the cuticle (leaf surfaces). The pathogens (e.g. fungi) then face a second barrier (the cell wall), which they degrade by secreting enzymes such as cutinases, pectinases, cellulases and polygalacturonases (Agrios, 2005). Finally, the virulence (pathogenicity degree) strategy of the pathogen depends on how it utilizes the plant cell nutrients. Biotrophs

(e.g. *Pseudomonas syringae*, *Ps*) obtain nutrients from living tissues while necrotrophs (e.g. *Botrytis cinerea*) feed on dead or dying cells. Some pathogens, referred to as hemi-biotrophs (e.g. *Phytophthora infestans*), can act both as biotrophs and necrotrophs, depending on the stage of their life cycle or the surrounding conditions (Glazebrook, 2005; Jones and Dangl, 2006).

1.2.1 Plant innate immunity

Plants, unlike mammals, do not have an adaptive immune system and defender cells that migrate to the source of infection and halt the danger. Instead they rely on the innate immunity of each cell and on systemic signals produced from infection sites (Dangl and Jones, 2001; Jones and Dangl, 2006). In general, plants show two types of responses upon invasion by a pathogen. They develop either disease (i.e., compatible interaction with the pathogen) or resistance, halting pathogen growth, also referred to as incompatible interaction (Katagiri et al., 2002). Innate immunity is the resistance to pathogens by triggering defense responses to terminate or restrict pathogen growth (Jones and Dangl, 2006). Innate immunity in plants can be developed through two approaches. First, through the recognition of pathogen-associated molecular patterns (PAMP) that limits pathogen infections and is referred to as PAMP-triggered immunity (PTI, 1.2.1.1). Second, the plant is able to induce defense responses after recognition of the so-called pathogen effectors, which is referred to as effector-triggered immunity [ETI, 1.2.1.2, (Jones and Dangl, 2006)].

1.2.1.1 PAMP-triggered immunity

Structural physical defenses, such as wax, cuticle on the leaf surfaces, and cell walls are the first obstacles to invading pathogens. Plants also have preformed chemical defenses that include antimicrobial compounds and secondary metabolites that can either be toxic to the pathogen or that can inactivate the enzymes secreted by the pathogen (Agrios, 2005). The recognition of PAMPs by plant pattern recognition receptors (PRRs) induces PTI, also referred to as basal resistance, and is considered as the primary plant immune response (Jones and Dangl, 2006). PAMPs generally contribute to a function that is critical to the organism life and thus, are indispensable and are generally well conserved across a wide range of microbes (Nürnberger and Kemmerling, 2009).

A major breakthrough in understanding PTI came when plants were found to recognize one specific PAMP, bacterial flagellin, the proteinaceous subunit that is the main component of the bacterial flagellum (Felix et al., 1999). Recognition of flagellin or a 22-amino acid peptide (flg22), derived from a well-conserved domain of flagellin, was found to inhibit growth of *Arabidopsis* seedlings, elicit callose deposition, trigger ROS and PR protein production, and trigger resistance to virulent bacteria in wild-type (wt) plants (Gomez-Gomez et al., 1999; Gomez-Gomez and Boller, 2000; Zipfel et al., 2004). The PRR that recognizes flg22 is the receptor-like kinase (FLS2, Figure 1.6) that initiates a signaling cascade through a mitogen-activated protein kinase (MAPK) and leads to the rapid transcriptional induction of a number of genes including WRKY transcription factors [Figure 1.6, (Asai et al., 2002)]. *fls2* plants, which have a non-functional FLS2, are not able to recognize flg22 (Zipfel et al., 2004; Heese et al., 2007). *fls2* plants are more susceptible to *Ps* when sprayed on the leaf surface than wt plants (Zipfel et al., 2004). In addition to flagellin, *Arabidopsis* has subsequently been shown to recognize several other bacterial PAMPs including bacterial elongation factor-Tu, the cell wall components peptidoglycan and lipopolysaccharides (LPS). The receptor mediating perception of elongation factor-Tu has been identified and is known as elongation factor-Tu receptor [EFR, Figure 1.6, (Nürnberger and Kemmerling, 2009)]. Fungal and oomycete PAMPs are mainly cell wall components such as chitin and β- glucan or lectin, respectively.

In general, PAMP recognition is followed by several physiological and molecular changes, for example Ca^{2+} fluxes are observed across the plasma membrane, and MAPK cascades are activated. Protein phosphorylation, callose deposition, cell wall thickening, stomatal closure and oxidative burst are also examples of PAMP-triggered responses (Nürnberger and Kemmerling, 2009). Indeed, the transcriptional response induced by different PAMPs not only shares many similarities, but also overlaps with ETI-mediated transcriptional changes (explained in 1.2.1.2) during incompatible interactions, indicating that PTI provides a broad-spectrum defense mechanism (Schwessinger and Zipfel, 2008; Nürnberger and Kemmerling, 2009).

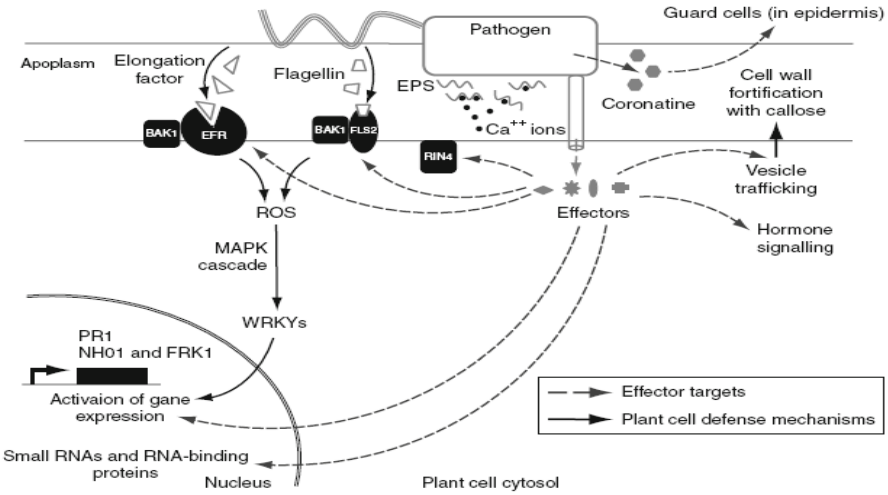


Figure 1.6: Effector targets to suppress immunity
The pathogen effectors and their targets are represented by filled gray symbols and connected by dotted lines. Mechanisms for PTI suppression include calcium chelation by extracellular polysaccharides, opening of stomata by coronatine, hormone signaling, blockage of vesicle trafficking. Plant defense mechanisms, PTI and ETI, are in black symbols and lines: Figure taken from (Metraux et al., 2009).

1.2.1.2 Effector-triggered immunity

Suppression of PTI is a major strategy of virulent pathogens to facilitate infection of susceptible host plants. The pathogen secrets effector proteins that inhibit plant major defense responses [Figure 1.6, (Metraux et al., 2009)]. Thereby, effector-triggered susceptibility is developed in the plant and might lead to disease, i.e., compatible interaction (Jones and Dangl, 2006). In addition to PAMPs, plants have evolved the ability to detect pathogen effectors, such as type III secretion system (TTSS) effectors directly secreted into the host cell by *Ps*, leading to ETI as a secondary line of resistance (Jones and Dangl, 2006; Heath, 2009). ETI is developed after recognition of specific type of effectors by host resistance (R) proteins (Figure 1.8), most of which belong to leucine-rich repeat (LRR) and nucleotide-binding site (NBS) domain containing proteins. When the effector protein is recognized by R protein, is called avirulence (avr) pathogenic determinant. In general, ETI induces signal transduction cascades that will lead to hypersensitive response (HR, explained in 1.2.2.2). HR involves death of the affected cell and is thought to be a form of programmed cell death (Heath, 2000; Heath, 2009). If either of avr or R genes is missing,

disease develops. This ETI phenomenon was firstly described as gene-for-gene resistance. It has been observed in a broad variety of pathogen infections, including bacteria, fungi, and viruses (Glazebrook et al., 1997; Jones and Dangl, 2006; Heath, 2009).

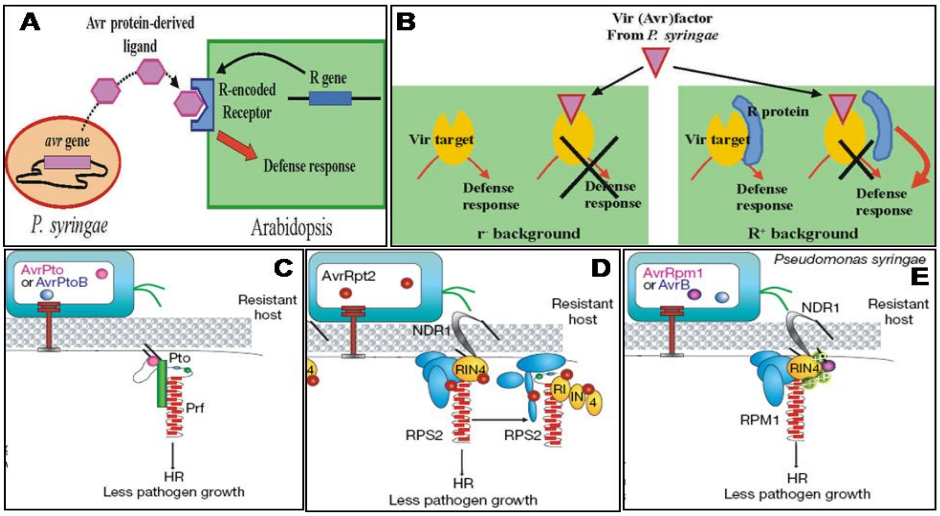


Figure 1.7: Molecular mechanisms of ETI initiation.
A: The ligand-receptor model of R and avr interaction. B: The guard model of R and avr interaction. When a plant does not have an appropriate R gene (left), an avirulence factor derived from *Ps* interacts with the virulence target. When a plant has the appropriate R gene (right), the virulence target is guarded by the R protein. C: Receptor-ligand model example. D and E: Guard model examples: Figures taken from (Katagiri et al., 2002; Jones and Dangl, 2006).

To explain the interaction of avr with R proteins, two different models have been proposed, the so-called ligand-receptor model and the guard model (Katagiri et al., 2002; Jones and Dangl, 2006). In fact, only few cases of ligand-receptor (Figure 1.7, A) explain direct interactions between avr and R proteins upon infection [e.g., the avrPto from *Ps* with tomato R protein Pto kinase (Figure 1.7, C)] (Tang et al., 1996). Instead, the guard model (Figure 1.7, B) proposes that R proteins guard the host targets of avr proteins. According to this model, any alteration of the host target by the effect of avr proteins will lead to an activation of R proteins that activate ETI, thus indirectly detecting the virulence effectors. The best characterized example of a guard model is the recognition of the host target, *Arabidopsis* RPM1-interacting protein 4 (RIN4). The *Ps* effectors avrB and avrRpm1 inactivate RIN4 by phosphorylation (Mackey et al., 2002). The phosphorylation of RIN4

leads to the activation of two *Arabidopsis* R proteins, resistance to *Ps* 2 (RPS2) and resistance to *Ps* pathover (pv.) maculicola 1 (RPM, Figure 1.7, E). Another example, the R protein RPS2 is activated by the absence of RIN4 caused by its proteolytic degradation (Mackey et al., 2003) by the *Ps* effector avrRpt2 (Figure 1.7, D) (Katagiri et al., 2002; Jones and Dangl, 2006).

R proteins share similar structures (Figure 1.8), suggesting common mechanisms in pathogen response, and are divided into five classes based on their structural motifs. One class are the NBS-LRR proteins that contain N-terminal NBS and C-terminal LRRs. The *Arabidopsis* genome contains 149 NBS-LRR-encoding genes from which two distinct groups of sequences were identified: those that encode an N-terminal domain with Toll/Interleukin-1 receptor homology (TIR-NBS-LRR), for example, *Arabidopsis* RPS4, and those that encode an N-terminal coiled-coil motif (CC-NBS-LRR), for example, RPS2 and RPM1 (Meyers et al., 2003). Based on studies in *Arabidopsis*, the two NBS-LRR subgroups employ different signalling pathways: TIR-NBS-LRR-mediated resistance is achieved through enhanced disease susceptibility 1 [EDS1, (Parker et al., 1996)], while CC-NBS-LRRs signal through non-race specific disease resistance1 [NDR1, see 1.2.3.1, (Century et al., 1997)].

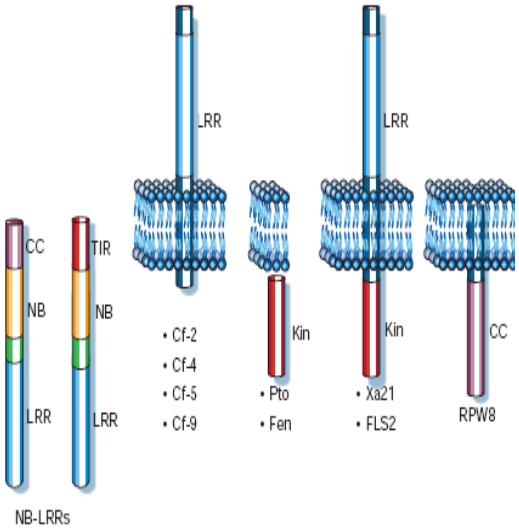


Figure 1.8: R protein representation
Location and structure representation for the five main classes of R proteins are presented. Xa21 and Cf-X proteins carry transmembrane domains and extracellular LRRs. RPW8 protein carries a putative signal anchor at the N terminus. The *Pto* gene encodes a cytoplasmic Ser/Thr kinase, but may be membrane associated through its N-terminal myristoylation site. The largest class of R proteins, the NB-LRR class, are presumably cytoplasmic (although they could be membrane associated) and carry distinct N-terminal domains: Figure taken from (Dangl and Jones, 2001).

1.2.2 Immune responses

1.2.2.1 Plant hormones

Plant hormones are implicated in diverse stress responses as well as developmental processes. Of these hormones, SA and JA, play major roles in modulating plant defense responses against various pathogens (Bari and Jones, 2009). In general, SA signaling mediates resistance to biotrophic and hemi-biotrophic pathogens, while the JA signaling pathway mediates resistance to necrotrophs and insects (Glazebrook, 2005). SA was reported to be important in systemic acquired resistance [SAR, explained in 1.2.2.3, (Gaffney et al., 1993; Delaney et al., 1994)]. Also, SA activates non-expressor of PR1 (NPR1) and triggers its translocation into the nucleus where it interacts with transcription factors that induce the expression of several defense-related genes including *PR*s (Shah, 2003). JA and its derivatives play important roles in plant development and physiology such as seed germination, fruit ripening, stomatal opening and root growth. They were first shown to play important roles in regulating wound- and insect-induced pathways (Creelman and Mullet, 1997; Thaler et al., 2002). JA also plays a role in mediating plant resistance against certain fungal and bacterial pathogens. For instance, exogenous application of JA induces the production of phenolics, nicotine and numerous other secondary compounds as well as defense-related compounds such as thionin in *Arabidopsis* (Creelman and Mullet, 1997).

Ethylene influences several developmental processes such as germination, fruit ripening and senescence, but is also involved in modulating defense responses. In fact, ethylene levels increase during early plant responses to pathogens, and exogenous application of ethylene enhances the expression of defense related genes (Dong, 1998). Recent studies indicate that other hormones such as abscisic acid (ABA), gibberellic acid, cytokinin and brassinosteroids are also implicated in plant defense signaling pathways but their role in plant defense is less well studied (Bari and Jones, 2009).

1.2.2.2 Hypersensitive response

As mentioned above (1.2.1.2), HR is a phenomenon associated with ETI. Several lines of evidence indicate that this cell death response is genetically programmed and not caused by pathogen-secreted toxins (Greenberg, 1997; Greenberg and Yao, 2004). HR is generally thought

to contribute to defense against biotrophic and hemi-biotrohphic pathogens. Not only does it play a role in directly limiting pathogen growth, but HR also is associated with the activation of SAR (1.2.2.3), which results in increased resistance to subsequent infections. The importance of HR in disease resistance therefore depends on the host-pathogen interaction [ETI, 1.2.1.2, (Greenberg, 1997; Greenberg and Yao, 2004)].

1.2.2.3 Systemic acquired resistance

SAR is a protective systemic broad spectrum defense and is induced following infections by necrotizing pathogens (i.e., pathogens that cause necrotic lesions due to disease symptoms or HR). Four stages were suggested for SAR induction, 1) SAR long-distance signals are produced [methyl-SA (MeSA), JA, lipids and constitutive-disease tesistance 1 (CDR1)] and may bind defective in induced resistance 1 protein (DIR1), a putative signal chaperone in the induced leaf, 2) The signals move from the induced leaf to distant tissues via the phloem, cell-to-cell, and/or by the volatile MeSA, 3), and are perceived by signal receptors that might include NPR1 and unknown receptors (Glazebrook et al., 1997; Xia et al., 2004; Vlot et al., 2008; Champigny and Cameron, 2009). In *npr1.1* plants that carry a single recessive mutation in *NPR1*, the SAR-responsive expression of other *PR* genes is abolished (Cao et al., 1994), 4) subsequent pathogen infection allows the distant leaves to respond in a resistant manner (Glazebrook et al., 1997; Vlot et al., 2008; Champigny and Cameron, 2009). An important role of SA in SAR was supported by the fact that exogenous application of SA or SA analogs [2,6- dichloroisonicotinic acid (INA) and benzothiodiazole] induces resistance against pathogens. Moreover, SAR collapsed and pathogen susceptibility increased during expression of salicylate hydroxylase (nahG) from *Pseudomonas putida* that converts SA to catechol (Gaffney et al., 1993; Delaney et al., 1994).

1.2.3 Defense-related proteins

1.2.3.1 NDR1/HIN1 like proteins (NHLs)

The defense-associated gene, harpin-induced gene 1 (HIN1) was isolated from tobacco and shown to be induced by flg22 and *Ps* that induce HR (Gopalan et al., 1996). One *Arabidopsis* homolog of tobacco HIN1 is NDR1 that is involved in gene-for-gene mediated resistance

mechanisms in response to attack by both bacterial (e.g., *Ps*) and fungal pathogens (Century et al., 1997). The glycosylphosphatidylinositol (GPI)-anchored NDR1 is PM associated and is an essential protein for the activation of two R proteins, RPS2 and RPM1 (Coppinger et al., 2004; Jones and Dangl, 2006). It has been demonstrated that NDR1, RPM1 and RPS2 are capable of interacting with RIN4 protein, and that the activation of disease resistance develops once RIN4 protein is altered by the action of the bacterial effectors [see 1.2.1.2 and Figure 1.7, D and E, (Mackey et al., 2002; Day et al., 2006)]. The interaction with RIN4 was suggested to regulate activation of disease resistance signaling following recognition of *Ps* in *Arabidospsis* (Day et al., 2006). However, the mode of action of NDR1 remains elusive, but its overexpression in *Arabidopsis* resulted in enhanced bacterial disease resistance (Coppinger et al., 2004; Day et al., 2006). A fast-neutron mutant in *Arabidopsis* (*ndr1-1*) was more susceptible to the avirulent *Ps* strains expressing the effectors avrB, avrRpt2, avrRpm1, or avrPphB (Century et al., 1995; Century et al., 1997).

Arabidopsis carries a large number (i.e., 45) of NDR1/HIN1-like (NHL) proteins. Most of the NHLs share three conserved unique motifs of unknown function (Zheng et al., 2004). Recently, many defense roles of NHL proteins in plant resistance responses were reported. For instance, *NHL2* overexpression in transgenic *Arabidopsis* plants resulted in elevated levels of *PR-1* expression and light-dependent "speck disease-like" symptoms in the leaves (Dormann et al., 2000). Similarly to NDR1, NHL3 (post-translationally modified by glycosylation) is PM associated, and its overexpression results in bacterial disease resistance in *Arabidopsis* to virulent *Ps* (Varet et al., 2003). Based on expression analysis, *NHL25* was proposed to be used as a marker for incompatible interactions (ETI, 1.2.1.2) with pathogens and possibly for HR development, where it was induced in-parallel or upstream of the pathway that is mediated by NDR1 or EDS1 [see 1.2.1.2, (Varet et al., 2002)]. On the other hand, *NHL3* can be induced by biotic and abiotic stresses and is altered by avirulent pathogens in ETI (Varet et al., 2002). *NHL3* and *NHL25* induction were reported also to be SA independent and dependent, respectively, which indicates that *NHL*s are induced by different pathways of defense mechanisms (Varet et al., 2002). The expression level of *NHL*s was also reported to be upregulated by pathogens including viruses [e.g., *NHL1*, *NHL2*, and *NHL10*, (Zheng et al., 2004)].

1.2.3.2 Immune-associated nucleotide-binding proteins (IANs)

Guanosine triphosphate (GTP)-binding proteins (GTPases) catalyze GTP hydrolysis, which is the key process in intracellular signal transduction (Scheffzek et al., 1998; Leipe et al., 2002). Recently, a new family of GTPases has been reported (in both vertebrate immune cells and plant cells) to be induced during antipathogen responses. This family was first discovered in plants after the isolation of a gene in *Arabidopsis* after infection with *Ps* pv. maculicola carrying a specific effector protein (avrRpt2), and designated as avrRpt2-induced gene (*AIG1*). *AIG1* expression was found to be induced by both virulent *Ps* and specifically by avirulent *Ps* that are inducing HR [see 1.2.1.2 and 1.2.2.2, (Reuber and Ausubel, 1996)]. Liu et al., (2008) suggested that AIG1 may mediate plant disease resistance through RPS2-dependent resistance signaling pathway in *Arabidopsis* (see 1.2.1.2). However, no further studies were reported for *AIG* genes.

This protein family has largely been studied in humans and has important functions in development of the immune system and the regulation of immune responses [e.g., T-cell homeostasis, (Cambot et al., 2002; Krucken et al., 2004; Schnell et al., 2006)]. The family members are now referred to as immune-associated nucleotide-binding proteins (IAN), also known as GTPase of immunity-associated proteins (GIMAP) (Wang and Li, 2009). Most of the *IAN* genes are clustered in both plant and vertebrate genomes, for instance, 12 *Arabidopsis IAN* (*AtIAN*) family members are located on chromosome 1 (Liu et al., 2008). All IAN proteins have specific, conserved domains: an AIG1 domain that consists of five motifs (G1–G5) for GTP-binding and a conserved hydrophobic box between G3 and G4 unique to AIG1-like proteins, and a coiled-coil motif (Krucken et al., 2004; Liu et al., 2008). Human IAN proteins are localized in diverse subcellular compartments such as the cytoplasm, ER, Golgi complex or mitochondria, which implies function modes of IAN mediating signaling pathways (Wang and Li, 2009). For AtIANs, no localization studies were reported.

1.3 Thesis goals

The physiological functions of plant peroxisomes are numerous (see 1.1.1) and some of these functions (e.g., stress-related functions, 1.1.1.3) are not adequately covered because of an incomplete knowledge of the complete peroxisomal proteome. To be able to investigate peroxisome functions in detail, one long-term goal is to discover the entire peroxisomal proteome. Based on the *Arabidopsis* full genome sequence (Arabidopsis genome initiative, 2000), a bioinformatics-based definition of peroxisome targeting signals, and analysis of experimental proteomics approaches were able to detect several putative peroxisomal proteins (see 1.1.4).

Many putative peroxisomal proteins were identified by PTS1 prediction models, and experimental *Arabidopsis* leaf peroxisome proteomics. The aim of this study was to experimentally validate several putative peroxisomal proteins and targeting signals. Furthermore, to investigate peroxisome functions in plant innate immunity (see 1.1.1.3 and 1.2) by screening *Arabidopsis* proteins for PTS1 proteins with a potential role in defense-related functions. Several defense-related proteins were investigated using experimental validation combined with expression analysis, and followed by initiation of functional studies for selected proteins. The thesis goals were divided into five main sub-points that were studied in the course of this study:

1- Experimental validation of machine learning approaches (see 1.1.4.1) including two prediction models that were recently developed and identified several putative PTS1 tripeptides and *Arabidopsis* PTS1 proteins. The validation of the predicted PTS1s and PTS1 proteins to be investigated using the *in vivo* subcellular targeting studies.

2- Experimental validation of peroxisome targeting for novel candidate proteins with a predicted role in detoxification (e.g., GR1, DHAR1, and GSTs, 1.1.1.2). Furthermore, to initiate functional analyses for the peroxisomal ASC-GSH cycle enzymes (GR1 and DHAR1) and the peroxisomal GSTT1 by producing heterologus protein expressions and knockout mutants.

3- Investigation of the peroxisome defense machinery. The function of peroxisomes in defense responses is poorly studied (see 1.1.1.3) because of the difficulties to identify low abundance and stress-inducible peroxisomal proteins (see 1.1.1.3). In this study, several predicted defense-related proteins shall be experimentally validated by *in vivo* subcellular targeting studies.

4- NHL protein family (see 1.2.3.1) investigations. By screening *Arabidopsis* proteins for PTS1 proteins, several family members were identified by PTS1 prediction models. *In vivo* subcellular targeting analyses to be applied for the predicted NHLs. Furthermore, to study selected proteins expression analyses, followed by initiation of their functional analyses.

5- AtIAN protein family (see 1.2.3.2) investigations. AtIAN12 was detected in the proteome of *Arabidopsis* leaf peroxisomes (Reumann, unpub. data). Experimental validation of this protein together with other two homologs from the same family shall be carried out. Additionally, to study their expression analyses, followed by initiation of functional analyses for the selected proteins.

2. Materials and Methods

2.1. Materials

2.1.1 Enzymes and commercial kits

Commercial kit	**Source**
Wizard® Plus SV Minipreps	Promega, USA
Illustra GFX PCR DNA and Gel Band Purification Kit	GE Healthcare, England
pGEM®-T Easy Vector System	Promega, USA
Quick-change Site-Directed Mutagenesis Kit	Stratagene, USA
Expand high fidelity PCR system	Roche, Germany
Real-Time PCR Master Mix with ROX	Primerdesign, England
First Strand cDNA Synthesis kit	Fermentas, Germany
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems, USA
Invisorb Spin-Plant DNA mini Kit	Invitek, Germany
RNeasy® Plant Mini Kit	Qiagen, Germany
Phire® Plant Direct PCR Kit	Finnzymes, Finland

2.1.2 Bacterial strains

2.1.2.1 *Escherichia coli* (*E. coli*)

JM109 (Sigma-Aldrich, USA): JM109 is a K strain bacterium that carries the *rec*A1 and *end*A1 mutations. The *rec*A1 aids in plasmid stability while *end*A1 provides high quality plasmid preparation. JM109 cells also contain an F' episome carrying Δ(*lac*Z)M15 for blue-white screening via α-complementation with the amino terminus of β-galactosidase. The strain was kindly provided by Dr. Ioannis Livieratos, MAICh, Greece. The strain was largely used for cloning and subcloning purposes (see 2.2.2.4).

BL21 (New England Biolabs, England): BL21 is an *E. coli* B F- *dcm ompT hsdS* (r_B- m_B-) *gal*. The strain was used for heterologous expression of proteins.

SG13009 [pREP4] (Qiagen, Germany): SG13009 strain is derived from K12 strain and is useful for the production of proteins that are expressed with pQE vectors (see 2.1.3).

2.1.2.2 *Agrobacterium tumefaciens*

Agrobacterium tumefaciens is a soil-dwelling bacterium that transforms normal plant cells into tumor-forming cells by inserting a piece of bacterial DNA (the transfer, or 'T' DNA) into the plant cell genome. The T-DNA, is flanked by left and right border sequences, and presents on a tumor inducing (Ti) plasmid.

GV3101 (pMP90): GV3101 carries a disarmed Ti plasmid that possesses the virulence genes needed for T-DNA transfer, but has no functional T-DNA region of its own. GV3101 grows at 28-30°C and is resistant to rifampicin, while the Ti plasmid is resistant to gentamicin. The strain was used in subcellular localization-mediated transformations (see 2.2.3.2), and kindly provided by Prof. Jianping Hu, MSU, USA.

ABI-1: ABI-1 is a derivative of GV3101 (pMP90RK) which possesses the RK2 replicase and the trf gene required for plasmid replication. ABI-1 is resistant to rifampicin, while the Ti plasmid is resistant to kanamycin. The strain was used in stable transgenic lines-dependent transformations (see 2.2.1.4), and kindly provided by Prof. Simon G Møller, CORE, Norway.

2.1.2.3 *Pseudomonas syringae (Ps)*

Ps is a rod shaped gram-negative bacterium with polar flagella. The bacterium is a plant pathogen that can infect a wide range of plant species and exists as over 50 different pv. *Ps* enters the host tissues and in a susceptible plant it multiplies to high population levels in intercellular spaces. Infected leaves show water-soaked patches, which become necrotic which may be surrounded by diffuse chlorosis. In resistant plants, *Ps* triggers HR (see 1.2.2.2) and in return fails to multiply to high population levels and causes no disease symptoms. The susceptible interaction between *Arabidopsis thaliana* and the *Ps* is used as a model for host-pathogen interaction [see ET 1.2.1.2, (Anzai et al., 2000)]. The *Ps* strains used in this study were kindly provided by Prof. Sheng Yang He, MSU, USA.

***Ps* pv. tomato (*Pst* DC3000):** *Pst* DC3000 is a virulent strain of *Pst* that obtained resistance to rifampicin by spontaneous mutant generation. The *Pst* DC3000 complete genome was sequenced by The Institute for Genome Research (TIGR).

***Pst* DC3000 (avrRpt2):** *Pst* DC3000 (avrRpt2) is an avirulent strain of *Pst* DC3000 expressing the effector protein (avrRpt2), which is naturally secreted by *Ps* TTSS. In this strain, the avrRpt2 gene has been introduced by the pDSK600 plasmid (spectinomycin resistant) after transformation of the virulent *Pst* DC3000 (Mudgett and Staskawicz, 1999). The avrRpt2 gene expression causes the virulent strains of *Pst* DC3000 to be avirulent on *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) and other ecotypes (e.g., Niederzenz-0, Nd-0) which contain RPS2 resistance gene [see 1.2.1.2, (Innes et al., 1993)].

2.1.3 Vectors

pCAT: pCAT is a pUC based vector harbouring the *Cauliflower mosaic virus* (CaMV) 35S promotor with a duplicated enhancer region and a 35S polyadenylation site. 35S promoter is a very strong constitutive promoter, causing high levels of gene expression and is one of the most widely used promoters. pCAT was used as a backbone for generating pCAT-YFP vector (Figure 1.9) which is expressing enhanced yellow fluorescent protein (EYFP), and the pCAT-CFP vector which is expressing enhanced cyan fluorescent protein (ECFP). pCAT-CFP vector has been used to generate marker vectors for peroxisomes and mitochondria. Peroxisomal marker vector (gMDH-CFP) was generated by subcloning of a fragment containing the PTS2 sequence of glyoxysomal malate dehydrogenase (gMDH) from cucumber (Kim and Smith, 1994). Mitochondrial marker vector (coxIV-CFP) was also generated by subcloning a fragment containing the transit sequence of the cytochrome C oxidase IV subunit [coxIV, (Hurt et al., 1985)], a mitochondrial marker protein from yeast (Fulda et al., 2002). The vectors pCAT-YFP, gMDH-CFP, and coxIV-CFP vectors were kindly provided by Prof. Martin Fulda, Germany (Fulda et al., 2002).

Furthermore, pCAT-YFP vector was modified to obtain 2 other restriction sites downstream NotI (i.e., SacI and SacII) to allow possible subcloning combinations instead of XbaI. This modification led to formation of pCAT-YFP-M, which was used mostly in this study for subcloning of the genes of interest (Ma et al., 2006; Ma and Reumann,

2008). Finally, one additional vector (AK1-CFP) was created to obtain pCAT-CFP with presence of the restriction sites similar to pCAT-YFP-M. pCAT-YFP-M and AK1-CFP are used for generating N-terminal fusions for proteins of interest with the EYFP and ECFP, respectively. It's important to know that EYFP is lacking the stop codon in both vectors. In contrast, another vector (NS-EYFP) was also used where stop codon was normally available, and is used for N-terminal fusions of EYFP by NcoI.

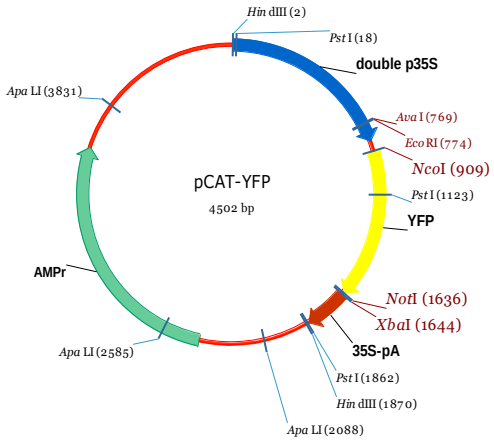


Figure 1.9: pCAT-EYFP vector map
pCAT-EYFP plasmid containing EYFP for transient expression in onion epidermal cells, and tobacco isolated protoplasts. The plasmid has a 35S promoter with a duplicated enhancer region and a 35S polyadenylation site, 35S-PA. The vector has been used in subcloning for the genes of interest in the back of EYFP.

Organelle markers: Several vectors expressing reporter fused proteins were obtained and used in this study. PWEN99 is a vector encoding red fluorescence protein-SKL (RFP-SKL) to label peroxisomes (Matre et al., 2009), kindly provided by Prof. Cathrine Lilo, CORE, Norway. A binary vector encoding CFP-SKL was also used to label peroxisomes (Zhang and Hu, 2008), kindly provided by Prof. Jianping Hu, MSU, USA. Moreover, a vector encoding orange fluorescence protein (OFP) fused with ER targeting signal, OFP-ER (Frank et al., 2008), kindly provided by Prof. Antje von Schaewen and Dr. Tanja Meyer, Germany. Finally, a set of binary vectors encoding CFP fused with targeting signals of ER, golgi, PM was also used (Nelson et al., 2007), kindly provided by Prof. Jianping Hu, MSU, USA.

pRS300: Vector used as a template for amiRNA (artificial microRNAs) construction (see 2.2.2.1). It contains the miR319a precursor in pBSK (cloned via SmaI site). To be able to generate amiRNAs, the amiRNA designer (WMD) delivers four oligonucleotide

sequences (I to IV) that were amplified from pRS300 and used to engineer amiRNA into the endogenous miR319a precursor by site-directed mutagenesis (Schwab et al., 2006). The vector was kindly provided by Prof. Cathrine Lilo, CORE, Norway.

pBA002: Binary vector contains CaMV 35S promoter, and confers resistance to spectinomycin and the herbicide glufosinate ammonium (alternative names: Basta, Phosphinotricin and Finale) in bacteria and plants, respectively. The genes which are responsible for resistanse against kanamycin and Basta are neomycin phosphotransferase and bialophos resistance gene (BAR) encoding phosphinotricin acetyl transferase enzyme, respectively. The vector was used in transient overexpression and in generation of transgenic stable lines (sees 2.2.1.4). The vector was kindly provided by Prof. Simon G Møller, CORE, Norway.

pER10.corReal: Binary vector contains Estradiol enhanced promoter, and confers resistance to spectinomycin and kanamycin in bacteria and plants, respectively. The vector was used in generation of transgenic stable lines (see 2.2.1.4). The vector was kindly provided by Prof. Simon G Møller, CORE, Norway.

pMAL-c2x: Vector designed to produce maltose-binding protein (MBP) fusions in *E. coli*, where the protein of interest can be cleaved from MBP with the specific protease factor Xa (New England Biolabs, England). The vector was kindly provided by Dr. Ioannis Livieratos, MAICh, Greece.

pQE31: Vector used to produce His_6-tagged proteins to be expressed in *E.coli*, and is based on the T5 promoter transcription-translation system (Qiagen, Germany).

2.1.4 Imaging facilities

2.1.4.1 Epifluorescence (Nikon)

Fluorescence image acquisition was performed on a Nikon *TE-2000U* inverted fluorescence microscope equipped with an Exfo X-cite 120 fluorescence illumination system (Exfo) and filters for YFP (exciter HQ500/20, emitter S535/30), CFP (exciter D436/20, emitter D480/40), a dual YFP/CFP filter with single-band exciters (Chroma Technologies), Texas red filter set for RFP: 31004, and chlorophyll autofluorescence (exciter HQ630/30, emitter HQ680/40, Chroma

Technologies, Rockingham, VT, USA). All images were captured using a Hamamatsu Orca ER 1394 cooled CCD camera. Volocity II software (Improvision, Coventry, UK) was used to capture 0.5 µm *Z*-sections to generate extended focus images.

2.1.4.2 Epifluorescence (Zeiss)

Epifluorescence microscopy was performed with an *Axio Imager M1* microscope (Carl Zeiss) for visualization of YFP labeled proteins (excitation 500 ± 12 nm; emission 542 ± 13.5 nm) and callose depositions using the 4,6-diamidino-2-phenylindole (DAPI) excitation filter (excitation 400/418 nm; emission 478/495 nm). Axiovision Rel.4.8 program was used to capture images. This facility was used at MSU/USA during the two months visit.

2.1.4.3 Confocal (Nikon)

A confocal laser-scanning microscope from Nikon *A1R* was used to obtain images of fluorophore labeled proteins. Laser beams used for fluorophore excitation were: CFP, 457 nm; YFP, 514 nm; and chlorophyll, 638 nm. For emission, the following filters were used: 475/500 nm band pass for CFP, 520/555 band pass for YFP, and 650 nm long pass for chlorophyll.

2.1.4.4 Confocal (Olympus)

A confocal laser-scanning microscope from Olympus "*Fluoview FV1000*" was used to obtain images of fluorophore labeled proteins transformed in tobacco leave cells. Laser beams used for fluorophore excitation were: CFP, 458 nm; YFP, 514 nm; MitoTracker red, 543 nm; and chlorophyll, 633 nm. For emission, the following filters were used: 475/500 nm band pass for CFP, 520/555 band pass for YFP, 560/614 band pass for Mito-Tracker, and 650 nm long pass for chlorophyll. All images were acquired from single optical sections. This facility was used at MSU/USA during the two months visit.

2.2. Methods

2.2.1 Plant material and growth conditions

2.2.1.1 *Arabidopsis* seed sterilization

Arabidopsis seed surface sterilization was carried out in a sterile flow cabinet. The seeds were soaked in 1 ml solution [70% (v/v) ethanol and 0.05% (v/v) Triton X-100] for 10 min with occasional shaking. The seeds were then washed twice in 100% ethanol for a total of 10 min and dried on a sterile filter paper. The seeds were next spread on the surface of 0.8% (w/v) agar plates containing 1% (w/v) sucrose and 1/2 Murashige and Skoog (MS) containing vitamins. The sown seeds were subsequently stratified at 4ºC in the dark for a period of 2 days before being transferred to standard growth conditions (see 2.2.1.2).

2.2.1.2 Standard growth conditions

For plants grown on soil, *Arabidopsis* seeds were sown on a mixture of commercial soil (P-jard, LOG/ Oslo, Norway) and Perlite (3:1) and grown at ~22°C with a light intensity of 100~150 μmol m^{-2} s^{-1} in a 16/8 h cycle (long-day). The soil was treated once weekly with Hoagland nutrient solution, if required (Hoagland and Arnon, 1950). After sowing the seeds, they were covered with a plastic dome for the first week to maintain humidity until germination.

2.2.1.3 Growth conditions for immune assays

Arabidopsis seeds were sown in soil and covered with a plastic dome to maintain high humidity for efficient germination. The growth chamber conditions are 22°C and 70-80% relative humidity with 12 h of fluorescent light (intensity of approximately 100~150 μmol m^{-2} s^{-1}). After a week, the plastic domes were removed. Plants 4 to 6 weeks old were used for bacteria inoculation [at this point they usually have numerous large leaves but have not started to flower, (Katagiri et al., 2002)].

2.2.1.4 Plant transformation

Transgenic *Arabidopsis thaliana* was generated using *Agrobacterium*-mediated transformation based on the floral dip method (Clough and Bent, 1998). *Arabidopsis* plants were grown in 10 cm^2 pots under standard growth conditions (see 2.2.1.2). The primary inflorescences were clipped to promote the generation of secondary inflorescences.

The plants were used for transformation when the secondary inflorescences had reached no more than 10 cm in height and had a few open flowers.

Luria-Bertani (LB) culture media (200 ml) of *Agrobacterium* containing the desired construct (see 2.2.2.1) supplemented with selectable markers, was grown at 28°C overnight until the cells reached early stationary phase. The cells were spun down and resuspended in about 200 ml of dipping solution (0.5% (w/v) sucrose and 10 mM $MgCl_2$) and 0.05% Silwet L-77 (Lehle Seeds, USA) added immediately prior to dipping. The inflorescences of *Arabidopsis* plants were then dipped into the *Agrobacterium* suspension for 10-20 min. Excess liquid was then gently shaken from the plants, and the plants were laid down and loosely covered with plastic cover to maintain a humid environment. 24 hr after dipping, the cover was removed and the plants then grown under standard conditions until the siliques were dry and the seeds were ready for harvesting. The seed bulk was harvested and the first generation was screened for transformants. Screening for T1 seeds was performed on MS agar plates containing 10 $\mu g.ml^{-1}$ Phosphinotricin (PPT) for plants transformed by pBA002 derived constructs (see 2.1.3) or 50 $\mu g.ml^{-1}$ kanamycin for plants transformed by pER10 derived constructs [see 2.1.3, (Weigel and Brook, 2002)]. Marker resistant seedlings were selected 10-14 days after germination and transferred to fresh plates before being transplanted to soil. T1 plants were screened to validate successful transformation by genotyping of genomic DNA of the primary transformants by primers upstream (forward) and downstream (reverse) of the cDNA insertion sites in the transformed vector.

2.2.1.5 Characterization of T-DNA insertion mutants

T-DNA insertion seeds (Table 2.1) were first grown on MS agar plates (see 2.2.1.1) and then transferred to soil after germination. Homozygous mutants were identified by PCR analysis of genomic DNA isolated by Phire plant PCR kit (see 2.1.1) using gene-specific forward (LP), T-DNA left border primers and a gene-specific reverse primer (RP). The LP and RP primers (Appendix, Table 2.4) were designed by T-DNA Primer Design tool; http://signal.salk.edu/tdnaprimers.2.html (Yan and Robert, 2008). Five mutant were kindly also obtained from Prof. Sheng Yang He and Prof. Jianping Hu, MSU (*fls2.17*, *npr1.1*, *ndr1.1*, *pen2-1* and *pen2.2*).

Table 2.1: T-DNA insertion lines
The lines were obtained from the *Arabidopsis* Biological resource center (ABRC, Ohio, USA). Successful number of homozygous mutants are indicated, and the location of the T-DNA in the gene.

AGI code	Gene	T-DNA lines	Hom. No.	Insertion location
At3g24170	*GR1*	SALK_105794C	4	Intron
At1g19570	*DHAR1*	SALK_005382.46.25.x	2	Exon
At5g41210	*GSTT1*	SALK_014245.39.15.x	--	Exon
At3g51660	*AtMIF1*	SAIL_892_D10	--	Intron
AT4G14930	*AtSurE*	SALK_037615	4	Intron
AT5G17890.1	*AtLIMDP*	SALK_024264	2	Exon
At1g54540	*NHL4*	SAIL_681_E12	3	300-UTR
At1g65690	*NHL6*	SALK_148523	6	Exon
At5g36970	*NHL25*	SALK_113216	4	Exon
At3g54200	*NHL39*	SAIL_204_E02	3	Exon
AT5G21130	*NHL13H1*	SALK_080000	2	Exon
At3g05975	*NHL39H1*	SAIL_1213_B03	--	Exon
At4g09930	*AtIAN11*	SAIL_404_H08	2	300-UTR

2.2.1.6 Tobacco (growth conditions)

Nicotiana tabacum cv. Petit Havana (used for protoplast isolation, 2.2.3.3) seeds were surface-disinfected with 70% (v/v) alcohol for 1 min, and subsequently by 25% (v/v) bleach for 15 min, followed by four rinses with autoclaved deionized water. For each washing step, seeds were centrifuged at 14,000 rpm, and the liquid was decanted. Seeds were placed into plates containing a medium consisting of 3% sucrose, 1 MS and solidified with 0.8% (w/v) plant agar that had been adjusted to pH 5.8. Germinating seeds were placed in the culture chamber under 12/12 h light cycle at 22°C. After 2-3 weeks, germinating seedlings were transferred to Magenta boxes containing the same media and placed under the same light and temperature conditions to allow further growth. *Nicotiana tabaccum* plants (used for *Agrobacterium*-dependent transient transformation) were planted on soil and incubated at long day (18h day "23^{o}C"/6h dark "18^{o}C") at 60-70 µEinsteins light. The low light incubation conditions were used in order to optimize the leaves for subcellular localization experiments (see 2.2.3.2). Available *Nicotiana tabaccum* plants were picked from the green house facility, PRL, MSU, USA.

2.2.2 Molecular biology methods

2.2.2.1 PCR

To study the subcellular targeting of *Arabidopsis thaliana* full length-cDNAs with predicted PTS1s, fusion proteins with N-terminally located EYFP were generated. *Arabidopsis* cDNAs were ordered from ABRC center (Ohio, USA) and the BioResource Center (RIKEN, Ibaraki, Japan, Appendix: Table 2.4) or amplified by RT-PCR from the plant isolated RNA (see 2.2.2.2). Moreover, single exon gene (*NHL13H1*) was amplified by PCR from isolated genomic DNA. The proofreading High Fidelity Expand Polymerase (see 2.1.1) was used to amplify DNA fragments with conditions suggested by the manufacturer. Primers containing appropriate restriction endonucleases (Appendix: Table 2.4) were used for the amplification and further subclonings into the plant expression vectors (see 2.1.3). For EYFP-fused peroxisomal terminal domain (PTD), the C-terminal 10 residues of plant full-length proteins were fused to the C-terminus of EYFP by PCR using extended reverse primers and subsequently subcloned into empty pCAT vector (see 2.1.3).

To generate overexpresser lines, gene specific primers (Appendix: Table 2.4) were used to amplify full length *Arabidopsis* cDNAs of *NHL4*, *NHL6*, *NHL25*, *AtIAN12*, *AtIAN11*, and *AtIAN8*. Additional N-terminally fused proteins (NHL4, NHL6, AtIAN12, AtIAN11) with EYFP were generated, after constructing an intermediate vector (pGEMT-EYFP) where the selected genes were subcloned in the back of EYFP. Subsequently, the available EYFP-fused and non-fused fragments were excised and subcloned into the binary vectors pBA002 and pER10 (see 2.1.3). The resulting constructs were transformed (see 2.2.1.4) into *A. tumefaciens* strain ABI-1 (see 2.1.2.2) via the freeze-thaw method (Holsters et al., 1978). The resulting constructs were transformed to the wt *Arabidopsis* Col-0 plants by the floral dip method (see 2.2.1.4).

To produce tagged (His_6 and MBP) recombinant proteins in *E.coli*, cDNAs of *GR1*, *DHAR1* and *GSTT1* were amplified using gene-specific flanking primers and subcloned into pQE31 and pMAL.c2X (see 2.1.3).

To generate loss-of-function lines for (NHL4, NHL6, NHL25, AtIAN12, AtIAN11, and AtIAN8) the Web MicroRNA Designer

platform (WMD) was used to design amiRNA sequences (21mers) based on their annotations. Two different 21mers (amiRNAs) were selected per target or two targets at once (AtIAN11+AtIAN12, Appendix: Table 2.4). Each primary amiRNA construct was engineered from pRS300 (see 2.1.3) by modified PCRs, in a similar way, as described earlier (Schwab et al., 2006). The plasmid information for pRS300 has been integrated into the online WMD2 platform, and all appropriate primer sequences, needed for customization of pRS300, can be retrieved using the primer design function of WMD2. For each amiRNA construct, three overlapping fragments including the multiple cloning sites (MCS) were PCR amplified from the template (pRS300) using a total of six primers (4 are amiRNA-specific, and 2 are vector-specific). The three resulting fragments were gel purified and then fused in a single PCR with the two vector-specific flanking primers (Appendix: Table 2.4). The final fusion product of 554 bp was again gel purified, cloned into pGEM®-T Easy Vector (Promega, USA). The obtained constructs were sequence verified, excised with XhoI/SpeI and transferred into the MCS of the binary vectors pBA002 and/or pER10 (see 2.1.3).

2.2.2.2 RT-PCR

Total RNA was extracted using Triazol (Invitrogen, USA), according to the manufacturer's protocol. First-strand cDNA synthesis was performed using Superscript III reverse transcriptase (Invitrogen, USA) in a 20-µl standard reaction containing gene-specific primers. *NHL25* and *NHL6* cDNAs were amplified by reverse transcription-polymerase chain reaction (RT-PCR) from SA-treated leaves as shown previously (Varet et al., 2002), and senescent leaves, respectively.

2.2.2.3 Agarose gel electrophoresis

Confirmation of PCR products or restriction digests was regularly processed by agarose gel electrophoresis. DNA samples were mixed with 5x loading buffer (Fermentas, Germany) and loaded into agarose gels mostly consisting of 1% (w/v) agarose melted in 1x TAE (40 mM Trisacetate and 1 mM EDTA, pH 8.0) and 1:10000 diluted SYBR® Safe (Invitrogen, USA) or ethidium bromide. A 1 kb ladder (Fermentas, Germany) was loaded next to the samples as a DNA size marker. Samples were separated by electrophoresis in 1x TAE buffer at 100 V and visualized under UV light.

2.2.2.4 Transformation of competent *E.coli* cells

Competent *E.coli* (JM109, 2.1.2.1) cells which were prepared as shown previously (Chung et al., 1989) were placed on ice to thaw. The target vectors are added to the cells and incubated for 20 min. The cells were then given a heat-shock at 42ºC for 50 s and returned to ice for 2 min before adding 500 µl of LB medium. The cells were then incubated at 37ºC for 1-2 h with constant shaking (200 rpm) to allow plasmid replication and expression of the antibiotic resistance gene. 200-400 µl of the competent cells were spread on LB agar plates containing the appropriate antibiotics, and left to dry before incubation at 37ºC overnight.

2.2.2.5 Colony PCR

Direct colony PCR was used to screen for successful plasmid transformation into *E. coli* or *A. tumefaciens* colonies. Even though blue/white screening was used sometimes to determine if inserts are present, but also this technique facilitates determination of insert size and/or orientation in the vector. The homemade thermostable DNA polymerase from *Thermus aquaticus* ("Taq DNA polymerase") was used together with 10X PCR buffer (500 mM KCl, 100 mM Tris-HCl (pH 9.0), 1.0% Triton X100) and 25 mM $MgCl_2$ to set up the PCR reaction. For *E.coli*, a small amount of a colony were added and mixed well with the PCR reaction, while 150 µl from a grown culture of *A. tumefaciens* were centrifuged and the pellet was resuspended in 20 µl of water, which were boiled for 10 min, centrifuged and 3-5 µl from the supernatant were added to the PCR reaction.

2.2.2.6 Sequencing

The new recombinant constructs were isolated from transformed bacteria using Wizard® Plus SV Minipreps (see 2.1.1). Sequencing was done by Seqlab (Goettingen, Germany) using their facility of Extended Hotshots reactions which were applied for all of the new recombinant constructs. The general promoter T7 and SP6 primers were used for sequencing of the cloned inserts in pGEM-T Easy plasmid. For pCAT cloned inserts, vector backbone primers were used for sequencing (Appendix: Table 2.4). Sequence analysis was done using Vector NTI (Invitrogen, USA) in combination with web based programs for reversing DNA (http://www.bioinformatics.org/SMS/rev_comp.html) and protein translation (http://us.expasy.org/tools/dna.html).

2.2.2.7 Site-directed mutagenesis (SDM)

SDM was carried out using the QuickChange® Site-Directed Mutagenesis Kit (see 2.1.1). Primers containing the desired mutations, Appendix: Table 2.4) were designed according to the manufacturer recommendations. The plasmid DNA template was amplified by PCR using the *PfuTurbo*® DNA polymerase. Next, the methylated template plasmid DNA was removed by digestion with DpnI (10 U, Fermentas, Germany) for 1 h at 37ºC. Following the incubation, the nicked vector DNA containing the desired mutations was then transformed into *E.coli* (XL1-Blue) super competent cells supplied by the manufacturer. The clones obtained were then sequenced to confirm the presence of the desired mutation. AtLIMDP and DHAR1 (domain) constructs were mutated using this method.

2.2.2.8 Real-time PCR

Plants were grown and treated either by elicitor or pathogen (see 2.2.4). The treated leaves were frozen in liquid nitrogen and stored at -80°C before being ground into powder using liquid nitrogen. Total RNA was isolated using RNeasy® Plant Mini Kit (see 2.1.1). RNA was quantified by NanoDrop 2000 (Thermo Fisher, USA) and the concentration was adjusted to 100 ng/µl. The High Capacity cDNA Archive Kit (see 2.1.1) was used, according to the manufacturer's recommendations, to synthesize cDNA (50 ng/µl) using 1 µg RNA, which was further diluted to 10 ng/µl. Real-time PCR reactions were assayed using an ABI 7300 Fast Real-Time PCR System (Applied Biosystems, USA) with Sybr-Green for detection. The standard reaction volume was 20 µl containing 10 µl qPCR Master Mix (PrimerDesign, England), 300 nM primer (each of forward and reverse, Table 2.2) and 10 ng cDNA. Standard cycling conditions (2 min at 50°C, 10 min at 95°C and 40 cycles altering between 15 s at 95°C and 1 min at 60°C) were used for product formation. Comparative C_T method was used for relative quantitation of gene expression. Gene expression for each sample was calculated on three analytical replicates normalized using the average of the reference gene *Actin2*, using water treated tissues as calibrator. Thus, relative quantity of any gene is given as fold change relative to the calibrator.

Primers optimization and testing of the genes in this study were carried out by Chimuka Mwaanga's master thesis (Mwaanga, 2011). It was concluded from his study that *NHL6, NHL25* and *PR2* are equal to

Actin2 in regards of the amplification efficiency, hence their primers could be used for relative quantitation of gene expression. In contrast, *NHL4*, *AtIAN8*, *AtIAN11*, and *AtIAN12* amplification efficiencies were different from *Actin2* and refers that they need further optimization or replacement of the primers used, and could affect negatively on relative quantitation of gene expression. Neverthless, because the time limitations all the genes tested were used in this study as a preliminary step of analyses, bearing in mind the possible changes in relative quantitation for NHL4 and AtIANs.

Table 2.2: Real-time PCR primers
Forward and reverse primers were designed by QuantPrime (http://www.quantprime.de). The primer optimization and testing were carried out by Chimuka Mwaanga's master thesis (Mwaanga, 2011).

AGI code	Gene	Forward primer 5'-3'	Reverse primer 5'-3'
At3g18780	ACT2	TGCCAATCTACGAGGGTTTC	CAGTAAGGTCACGTCCAGCA
At1g54540	NHL4	TGCAGCAGCAACAACAAACAGG	TTCCGAGTTTGATGGCGACAGG
At1g65690	NHL6	TGGGAGCAAGATTACCGTGTGG	TTTGGCAACGACCCATTGCTTAG
At5g36970	NHL25	CCAGAATCAGTAATGGGTCGTTGC	CCTGTTAACCGTTGTTGCTCTTGC
At4g09940	IAN12	AGAGTTCAACGCTACCCAATGGC	TGGCGACAGACTAAACAGACCAG
At4g09930	IAN11	TGGCCAAGAAGGTAGAGAAGGTG	TCTTCGCTGGATTCTTCGTGGAG
At1g33960	IAN8	TCAATGTGATTGACACTCCTGGTC	ACTAAGAGCACAGCGTGTAGCC
At3g57260	PR2	AGCTTCCTTCTTCAACCACACAGC	TGGCAAGGTATCGCCTAGCATC

2.2.3 *In vivo* subcellular localization analysis

2.2.3.1 Transient expression in onion epidermal cells

Five micrograms of recombinant genes (see 2.2.2.1) were precipitated on 1.0 mm gold particles. Onions were cut into pieces and placed on a wet tissue in Petri dishes. These whole pieces were bombarded using a Biolistic Particle Delivery System (BioRad, USA) with 1100 psi rupture discs (briefly rinsed by ethanol) under a vacuum of 0.1 bars. After bombardment the samples were placed on a benchtop for 20 h in the dark. Onion epidermal cell layers were peeled and transferred to glass slides for microscopy (Fulda et al., 2002). The onion epidermal cell layer could be further incubated at 4^{o}C up to 8 days while keeping the sample humid. The longer incubation at cold temperature, allowed detection of weakly targeting proteins (Lingner et al., 2011).

2.2.3.2 Transient expression in tobacco leaves by *Agrobacteria*

Four to six weeks-old *Nicotiana tabacum* plants (see 2.2.1.6) were used for the *Agrobacterium tumefaciens*-mediated transient expression assays. *A. tumefaciens* strain GV3101(pMP90) (see 2.1.2.2) containing the recombinant genes was allowed to grow at 28^{o}C overnight, washed, and resuspended in water to an optical density at 600 nm of 0.5. Cells transformed with plasmids harboring either the EYFP fusion or organelles markers (see 2.1.3) were mixed and infiltrated into tobacco leaves using 1 ml needleless syringes. Leaves of infiltrated plants were analyzed after 2 days (Reumann et al., 2009).

2.2.3.3 Transient expression in isolated protoplasts

Tobacco protoplasts were transfected by a method described previously with minor modifications. Solutions used for the isolation and transformation are described at Table 2.3. Briefly, 3-4 leaves of 4-6 week-old *Nicotiana tabacum* cv. Petit Havana, grown in magenta boxes (see 2.2.1.6), were cut into small stripes with a sharp-razor blade and incubated with 12 ml enzyme solution at 28^{o}C for 16 h. After incubation, the protoplast suspension was filtered through two mesh sizes (125 µm and 63 µm) and protoplasts were collected by centrifugation at 60 g for 5 min. The pelleted protoplasts were resuspended in 10 ml W5 solution, incubated for 1 h on ice, and centrifuged. At this step, Haemacytometer slide was used to count the total number of protoplasts obtained.

To transform DNA into protoplasts, protoplasts were pelleted again and resuspended in MaMg solution, bearing in mind that the final protoplast number should be adjusted to 0.5 million/300 µl MaMg solution. Plasmid DNA (5–30 µg) was added to 300 µl MaMg solution containing protoplasts followed by 500 µl PEG solution. The mixture was incubated for 30 min at RT. After incubation, the mixture was centrifuged and the protoplasts were recovered in 3 ml B5 solution and incubated at RT in the dark. The expression of proteins was examined at various time points after transformation, potentially after 24 h and 48 h (Meyer et al., 2011).

Table 2.3: Solutions for protoplast preparation

Solution	Contents concentrations	Volume	Weights	Sterilization
Enzyme	0.5 M Mannitol for tobacco 10 mM CaCl2x2H2O 1% Cellulase (Onozuka R-10, Japan) 0.25% Macerozyme (Onozuka R-10)	100 ml	9.109g 0.147g 1g 0.25g	Sterile filter Freeze (12 ml-aliquots) at -20°C.
Mannitol	(0.5 M for tobacco)	500 ml	45.542g	Autoclave
CaCl2	0.2 M CaCl2x2H2O	250 ml	7.35g	Autoclave
W5	145 mM NaCl 125 mM CaCl2x2H2O 5 mM KCl 5 mM Glucose Check pH (5-6), or adjust	500 ml	4.237g 9.188g 0.186g 0.450g	Autoclave
MaMg	0.5 M Mannitol 15 mM MgCl2 x6H2O 0.1% MES Adjust pH (5.7) with 0.1 N KOH	50 ml	4.555g 0.152g 0.1g	Sterile filter
PEG	0.4 M Mannitol 0.1 M Ca(NO3)2 x4H2O 0.1% MES Adjust pH (8) or (7-9) with NaOH Poly ethylene glycol (PEG) 6000	100 ml	7.3g 2.362g 0.1g 40 g	Sterile filter Freeze at -20°C
B5	3.17 g/l Gamborg (Duchefa, Netherland) 0.5 M Glucose Adjust pH (5.7) with 0.1 N KOH	500 ml	1.585g 45.04g	Sterile filter Freeze at -20°C)

2.2.4 Immunity assays

2.2.4.1 Bacterial proliferation

Measuring bacterial multiplication within the host tissue is a method used to examine the plant innate immunity (see 1.2.1). A standard enumeration procedure involves pathogen inoculation followed by assaying bacterial populations present within host tissues at regular intervals. Bacteria used in this study were *Pst* DC3000 and *Pst* DC3000 (avrRpt2) (see 2.1.2.3). Bacteria were grown in low salt medium; LM (10 g l^{-1} Bacto tryptone, 6 g l^{-1} yeast extract, 1.5 g l^{-1} K_2HPO4, 0.6 g l^{-1} NaCl, and 0.4 g l^{-1} MgSO4.7H2O) with appropriate antibiotics (2.1.2.3). Virulent and avirulent *Pst* DC3000 bacteria were grown to the mid-logarithmic phase, centrifuged at 3000x g, and resuspended in a sterile water to the specified inoculums density. Syringe injections with relatively low inoculum densities ($1x10^6$ colony-forming units (CFU)/ml) were used. Four to six weeks-old *Arabidopsis* leaves (see 2.2.1.3) were infiltrated by pressuring bacterial suspensions into the

apoplast using a needless syringe. As a wounding control, distilled water was infiltrated into plant leaves. After inoculation, plants were left uncovered until leaves were no longer water soaked, then covered with humidity domes until completion of experiments (Gopalan et al., 1996; Katagiri et al., 2002). Leaves were harvested and leaf disks (0.38 cm^2) were excised from leaves with a cork borer number 5. The leaf disk for a single sample was placed in a 1.5 ml tube with 10 µl sterile distilled water, and ground with a plastic pestle by a small hand-held electric homogenizer. The pestle was then rinsed with 90 µl of water, with the rinse being collected in the original sample tube (total volume= 100 µl). A 10 µl sample was removed and diluted in 90 µl sterile distilled water. A serial 1:10 dilution series (up to 10^{-6}) was created for each sample. The diluted samples were placed on LM plates containing antibiotics, by spotting triple 10 µl aliquots of each of the serial dilutions and allowed to dry onto the surface. The plates were placed at 28°C for approximately 2 days; afterwards the CFU for each dilution of each sample are counted.

For the 10 µl spotting technique, a single spot was used for estimating the bacterial population only if it has >7 or < ~70 colonies. Plotting log (bacterial number/cm^2 leaf tissue) against time (in days), after pathogen inoculation produced the growth curve. Generally, this is a standard means of evaluating how well a bacterial pathogen multiplies in plant tissues (Gopalan et al., 1996; Katagiri et al., 2002). More than three bacterial number/cm^2 leaf tissues were averaged for determination of the CFU for each type of plants. Subsequently, standard deviation (SD) was calculated based on the difference of average numbers between bioliogical replicates, which in this case are two replicates (n=2).

2.2.4.2 Callose deposition

Callose (see 1.2.1.1), an amorphous, high-molecular-weight β-1,3-glucan is deposited in cell wall appositions (papillae) that form beneath infection sites and are thought to provide a physical barrier to pathogen penetration (Gomez-Gomez et al., 1999; Nishimura et al., 2003; Luna et al., 2011). By screening different ecotypes of *Arabidopsis* only wassilewskija (WS-0) was completely insensitive to the flagellin peptides (Gomez-Gomez et al., 1999).

Seeds of *Arabidopsis thaliana (*approximately 15 seeds per well) were planted in a sterile 12-well plate, each containing 1 ml filter-sterilized basal MS medium without Gamborg vitamins (Invitrogen, USA) with

0.5% (w/v) sucrose. Plates were kept in the dark at 4°C for 1–2 days for stratification before transferring them to the controlled growth cabinets. Seedlings were cultivated under standard growth conditions (see 2.2.1.2) but continuous light. After 8 days of growth, MS medium was replaced with fresh medium. At day 9, seedlings were treated with 1 μM flg22. This optimal flg22 concentration was based on previously reported dose-response experiments (Gomez-Gomez et al., 1999).

After another 24 h, seedlings were cleared and dehydrated with 100% ethanol. Seedlings were fixed in an acetic acid: ethanol (1:3) solution for 2 h and sequentially incubated for 15 min in 75% ethanol, next in 50% ethanol, and finally in 150 mM phosphate buffer, pH 8.0. Then they were stained for 1 h at 25^{o}C in 150 mM phosphate buffer, pH 8.0, containing 0.01% (w/v) aniline blue. After staining, seedlings were mounted in 50% glycerol. About eight leaves, from at least five independent seedlings were examined by UV epifluorescence microscope (see 2.1.4.2). Callose quantification was performed by using ImageJ software (Galletti et al., 2008). Five images representing 5 leaves from 5 independent plants were used for counting callose depositions, and their numbers were averaged, subsequently, SD was calculated based on the difference of average numbers between bioliogical replicates, which in this case are two replicates (n=2) containing 3 experiments.

2.2.5 Metabolic peroxisome function assays

2.2.5.1 Sucrose dependence

Arabidopsis and other oilseed plants β-oxidize long chain fatty acids in peroxisomes to provide energy during germination (see 1.1.1.1). Some mutants seeds germinate normally, but plants do not develop beyond germination unless provided with exogenous sucrose; a phenotype which suggests severe peroxisomal defects, because peroxisomal β-oxidation mutants cannot catabolize stored fatty acids for energy before photosynthesis begin (Hayashi et al., 1998). To determine whether disruption of a gene in a specific mutant will lead to impaired seedling establishment, hypocotyls lengths of dark-grown seedlings germinated in the presence or absence of sucrose should be tested (Zhang et al., 2010).

Seeds of wt *Arabidopsis thaliana* and mutants were sown on ½ Linsmaier & Skoog with vitamins (LS; caissonlabs, USA) agar growth

medium with or without 1% (w/v) sucrose, and stratified in the dark at 4°C for 2–4 days. Afterwards, seeds were allowed to germinate and grow in normal growth conditions (see 2.2.1.2) but in the dark for 5 days. Five-day-old etiolated seedlings were scanned using an EPSON scanner (http://www.epson.com). Hypocotyl length was then measured using IamgJ (http://rsb.info.nih.gov/ij/). More than 50 seedlings of each genotype were used for hypocotyl length measurements in three biological replicates (Zhang and Hu, 2009). Ten to 15 seedling hypocotyl lengths were measured and averaged, subsequently, SD was calculated based on the difference of average numbers between bioliogical replicates, which in this case are three replicates (n=3).

2.2.5.2 Auxin response

Indole-3-acetic acid (IAA) is a predominant auxin can be controlled in plants by altering rates of synthesis and degradation. Indole-3-butyric acid (IBA) is a second endogenous auxin; genetic evidence indicates that IBA is converted to IAA in peroxisomes. Because the conversion shortens the IBA side chain by two carbons, this process has been proposed to occur similarly to fatty acid β-oxidation. A collection of *Arabidopsis* mutants that are resistant to the inhibitory effects of IBA on root elongation but that respond normally to IAA were described and are mostly distinguished by developmental defects in the absence of exogenous sucrose, suggesting defects in peroxisomal β-oxidation (Zolman et al., 2001; Woodward and Bartel, 2005).

To study the response to IBA (final concentration 0, 10, 20, and 30 mM) was added to ½ LS agar growth medium with 0.5% (w/v) sucrose. Seeds from wt *Arabidopsis thaliana* and mutants were sown, followed by 2 days of cold treatment. To measure root elongations, seedlings were grown for 8 d under standard growth conditions (see 2.2.1.2) and the length of the primary roots was scanned using an EPSON scanner and measured using ImageJ (Zolman et al., 2001; Zhang and Hu, 2010). Ten to 15 seedling root lengths were measured and averaged, subsequently, SD was calculated based on the difference of average numbers between bioliogical replicates, which in this case are two replicates (n=2).

2.2.5.3 Photorespiration

During photorespiration process, peroxisomes are involved (see 1.1.1.1). In peroxisome defective mutants, the photorespiration could

be affected. Seeds of wt *Arabidopsis thaliana* and mutants were sown on ½ LS agar growth medium with or without 1% (w/v) sucrose, following 2 days of cold treatment. They were allowed to grow under standard growth conditions (see 2.2.1.2). When they are 2 weeks-old, plants were transferred to a freshly made soil and were allowed to grow for 18 days in a growth chamber under standard growth conditions (see 2.2.1.2) or low CO_2 (80 ppm).

2.2.6 Protein chemistry (SDS-PAGE)

SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was used to detect the overexpressed proteins. The recombinant vectors (see 2.2.2.1) were transformed to *E.coli* (see 2.1.2.1). The tagged proteins were expressed in *E.coli* by IPTG induction. The protein samples were mixed with 1x SDS loading buffer [60 mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS, 5% (v/v) β-mercaptoethanol, 0.025% bromophenol blue] were boiled at 100ºC for 5 min and separated according to their size. Unstained protein marker (Fermentas, Germany) was run alongside the samples and used as a size reference. SDS-PAGE gels consist of an upper stacking gel and lower separating gel. The stacking gel [125 mM Tris-HCl pH 6.8, 4% (w/v) acrylamide, 0.1% (w/v) SDS, 0.05% (w/v) APS, 0.15% (v/v) TEMED] was used for loading and concentrating the protein samples. A 10% SDS-PAGE separating gel [0.38 M Tris-HCl pH 8.8, 10% (w/v) acrylamide 0.1% (w/v) SDS, 0.05% (w/v) APS, 0.07% TEMED] fractionates proteins according to their molecular weight. The gels were fitted in a Mini-PROTEAN II cassette (BioRad) filled with SDS running buffer [250 mM Tris-HCl, 192 mM glycine and 0.1% (w/v) SDS]. Proteins were first electrophoresed at 80 V until they reached the end of the stacking gel, after which the voltage was increased to 150 V.

2.2.7 Leaf peroxisomes isolations

Four to six weeks-old *Arabidopsis* plant leaves (see 2.2.1.2) were harvested and leaf peroxisomes (n=5) were isolated as described previously (Reumann et al., 2007). These preparations will be used to study the ASC-GSH cycle (see 1.1.1.2) enzymes.

3. Results

3. 1. Validation of prediction models

3.1.1 *In vivo* validation of PTS1 tripeptides

Proteins are imported into peroxisomes mostly by a PTS1 or PTS2 [see 1.1.3, (Purdue and Lazarow, 2001)]. More than 100 new candidate proteins from plant peroxisomes had been identified, including low-abundance proteins, by both prediction models and proteome analyses ((Reumann, 2011) see 1.1.4). To better investigate the biological functions of peroxisomes, it is essential to identify the entire peroxisomal proteome. The prediction of plant peroxisomal proteins from genome sequences is an essential approach to identify additional yet unknown peroxisomal proteins (Reumann, 2011). A large data set (manuscript 1, Figure 1) of more than 2500 homologous plant sequences was generated from EST databases and 60 known *Arabidopsis* PTS1 proteins. Two prediction methods were applied to plant PTS1 proteins predictions: position-specific weight matrices (PWM) and residue interdependence (RI) models. Experimental verification supported the accuracy of both prediction methods (PWM and RI) on example sequences and identified several novel PTS1 tripeptides even including novel residues (manuscript 1, Table 1). Furthermore, several *Arabidopsis* proteins were predicted by PWM and RI models (see manuscript 1, Figure 4, and Supplemental data set 2).

The proposed PTD of the translated ESTs or proteins were N-terminally fused with EYFP (see 2.2.2.1), and their cDNAs were transiently expressed from the CaMV 35S promoter in onion epidermal cells that had been biolistically transformed (Fulda et al., 2002). Some plant sequences terminating with minor PTS1 tripeptides had already been predicted from 2004 dataset [SRV>, SML>, SNM>, etc., manuscript 1, Table 1, (Reumann, 2004)]. From this dataset, SRV> of the acyl-CoA oxidase 4 homolog of *Zinnia elegans* was validated as a functional PTS1, by detecting its EYFP-PTD in peroxisomes. However, organelle targeting of this construct could not be resolved under standard conditions (18 to 24 h at room temperature) but required extended expression times up to 1 week at reduced temperature (~ 10 ^{o}C). Indeed, the combination of cold incubation with the extension of expression time (from 24 h to 1 week) improved the detection

sensitivity for several weak targeting signals. The specificity of PTS1 protein import into peroxisomes was verified by EYFP alone and a few non-peroxisomal constructs (e.g., LCR> and LNL>), all of which remained in the cytosol under the same conditions (see manuscript 1). To further confirm SRV> as a plant peroxisomal PTS1, peroxisome targeting was validated for two additional PTDs of AGT homologs (SRV>). Both of their EYFP-PTDs were detected in peroxisomes. The targeting efficiency of both reporter fusions was different: SRV> (*Populus trichocarpa* x *Populus deltoides*) was weaker than the one from *Pinus taeda* (for more details see manuscript 1).

In the same study, the large data set was separated into three subsets (manuscript 1, Figure 1) based on the number of sequences that shared the same C-terminal tripeptide (1^{st}: most reliable data [≥3 sequences]; 2^{nd} [=2 sequences] and [3^{rd} =1 sequence]: uncertain data). From the 1^{st} data set, sixteen (e.g., CKI> and STI>) out of 42 identified C-terminal tripeptides had not been proposed to function as targeting signals by previous studies. Experimentally, CKI> and STI> were validated as novel functional PTS1 tripeptides (for more details see manuscript 1). To test the new algorithms for their ability to predict new PTS1 tripeptides, they were applied on the 2^{nd} and 3^{rd} uncertain data sets (manuscript 1, Figure 1). Several example sequences were selected for experimental verification based on their PWM and RI model-based prediction scores. Out of 12 example sequences chosen for experimental validation as part of the present dissertation, peroxisome and organelle targeting was validated for STI>, SPL>, PKI>, TRL>, and LKL> although with different efficiencies. Thus, these analyses identified five additional novel PTS1 tripeptides (STI>, SPL>, PKI>, TRL>, and LKL>). These results also added novel residues, namely Thr and Leu (position -3) and Pro, Phe, and Gln (position -2) to the plant PTS1 tripeptide motif ([TL][PFQ]z>). On the other hand, two other constructs (SGI> and SEM>) remained cytosolic. These results supported the assumption that these two uncertain data subsets are less reliable (for more details see manuscript 1).

The PWM and RI models were applied to the *Arabidopsis* genome using the gene model predictions of TAIR10. Out of the list of *Arabidopsis* genes which was provided based on their peroxisome targeting probabilities, 392 proteins (1.1% of the genome) were predicted to contain a PTS1. Approximately 271 gene models out of them had not yet been associated with peroxisomes. Experimentally,

EYFP-PTD of 1-aminocyclopropane-1-carboxylate synthase like pseudogene (ACS3, SPL>) was targeted to peroxisomes. Finally, several *Arabidopsis* full-length proteins (manuscript 1, Supplemental Table 5) were fused with EYFP (by the bachelor thesis, (Nilssen, 2009)) to investigate peroxisome targeting. The full-length Cys protease (SKL>) was targeted to peroxisomes, a Ser carboxypeptidase S28 family protein (S28FP, SSM>) directed EYFP to unknown subcellular vesicle-like structures, Nudix hydrolase homolog 19 (NUDT19, SSL>) was targeted to peroxisomes with lower effienency, and PfkB-type carbohydrate kinase family protein (pxPfkB, SML>) was also verified as a peroxisomal protein. Only a single full-length protein tested remained cytosolic (CUT1, VKL>, for more details see manuscript 1).

3.1.2 *In vivo* validation of PTS1 proteins

Investigation of peroxisomal targeting of predicted proteins was also investigated by extension of EYFP C-terminally by four additional predicted PTDs of constitutive triple response 1 (At5g03730.1/2, CTR1, SDL>), a self-incompatibility protein S1 family homolog (At2g23142, SPL>), an invertase/pectin methylesterase inhibitor superfamily homolog (At5g51500, K17N15.5, SEL>) and an FBD-like domain family protein (At5g53592, VKM>). All fusion proteins remained in the cytosol, except for SDL> which was verified to be in peroxisomes after extended incubation in cold (1 week) upon transient expression in onion epidermal cells (data not shown).

In order to improve efficient identification of putatively orthologous sequences, three *Arabidopsis* proteins that carried atypical PTS1 tripeptides, and preferentially represented low-abundance proteins were selected for experimental validation. These *Arabidopsis* proteins were fused in the back of EYFP [small thioesterase (sT4, SNL>, At1g04290), and two unknown proteins (At1g73970, UP10, ARL>; At4g33925, UP11, SKI>), and were validated in peroxisomes upon transient expression in onion epidermal cells (data not shown).

3. 2. Detoxification-related proteins

3.2.1 *In vivo* subcellular localization of detoxification proteins

Plant peroxisomes play essential roles in the detoxification of H_2O_2 through catalase and the ASC-GSH cycle (see 1.1.1.2). Peroxisomal GR and DHAR isoforms were identified by proteome analyses of *Arabidopsis* leaf peroxisomes, i.e., GR1 (At3g24170) and DHAR1 (At1g19570) (Reumann et al., 2007; Reumann et al., 2009). In the present study, it was found that *Arabidopsis* GR1 carries a novel PTS1-like tripeptide, TNL>, which had not been described as a plant PTS1 before. The residue T was also identified at pos. -3 in the PTS1 motif (see 1.1.4.1), which was not previously shown. Peroxisomal targeting for the EYFP-PTD (TNL>) was validated in both onion epidermal cells and tobacco protoplasts (see manuscript 2). The full-length GR1 was fused N-terminally with EYFP, and the fusion protein was detected in peroxisomes upon transient expression in onion epidermal cells, but not in tobacco protoplasts (see manuscript2).

DHAR1 was reported to be targeted to peroxisomes, when the full-length protein was fused C-terminally with EYFP and transiently expressed in intact tobacco leaves (Reumann et al., 2009). To investigate the PTS of DHAR1, the full-length DHAR1 was fused N-terminally with EYFP. The fusion protein remained in the cytosol upon transient expression in onion epidermal cells and tobacco protoplasts (Figure 3.1, A and J, 18-48 h expression time), indicating that the protein contains a PTS2 or an internal PTS rather than a PTS1. Interestingly, DHAR1 was found to contain a conserved PTS2-like domain ($RAx_{13}HL$) in the N-terminal domain (position 25 to 41, Figure 3.2). This peptide resembles PTS2 nonapeptides (e.g., $R[TMAV]x_5HL$) with the difference that the four conserved residues are spaces by 13 rather than five residues (Reumann, 2004). To investigate its subcellular targeting activity, the N-terminal domain of DHAR1 (46 aa) including the possible PTS2 domain was fused C-terminally with EYFP. Upon transient expression in onion epidermal cells, the fusion protein was indeed targeted to organelle-like structures (Figure 3.1, B). To investigate further whether the predicted atypical PTS2-like peptide directed the fusion protein (Nt_{46}-EYFP) to peroxisomes, SDM was applied to change the invariable residue, arginine, of the possible PTS2-peptide [($\underline{R}Ax_{13}HL$) to ($\underline{L}Ax_{13}HL$)]. In onion epidermal cells, the

point mutation did not abolish organelle targeting (Figure 3.1, C), indicating that this peptide did not act as a PTS2.

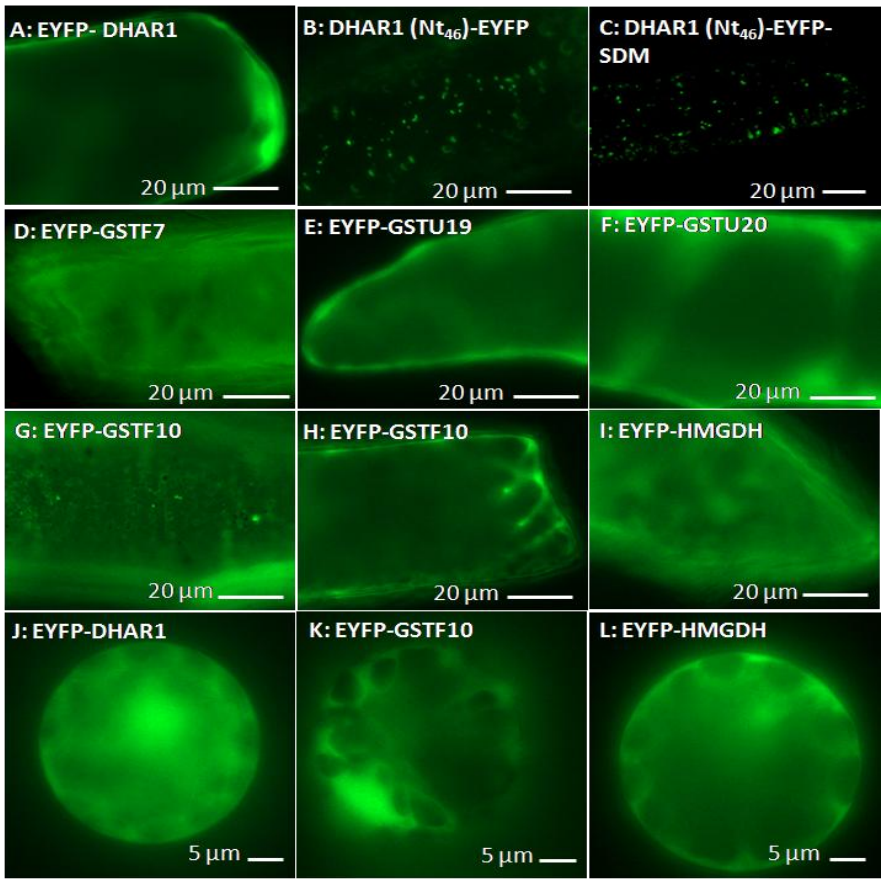


Figure 3.1: *In vivo* subcellular localization of DHAR1 and GSTs
A and D-L: The full-length proteins (DHAR1, GSTF7, GSTU19, GSTU20, and HMGDH) were fused N-terminally with EYFP and transiently expressed in onion epidermal cells or tobacco protoplasts. Apart from GSTF10, all fusion proteins remained cytosolic. GSTF10 mostly remained in the cytosol (H), but sometimes was targeted weakly to unidentified organelle-like structures (G). B and C are C-terminally fused DHAR1 (N-terminal 46 aa) with EYFP, and the SDM (R to L) of the domain construct containing (RAx13HL), respectively. Both of the EYFP-DHAR1 domains localized to organelle-like structures. J-L are images representing cytosolic targeting of DHAR1, GSTF10, and HMGDH fusion proteins in tobacco protoplasts. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown, except for J and C (n=2) and K and L (n=1). Expression times (18-48 h).

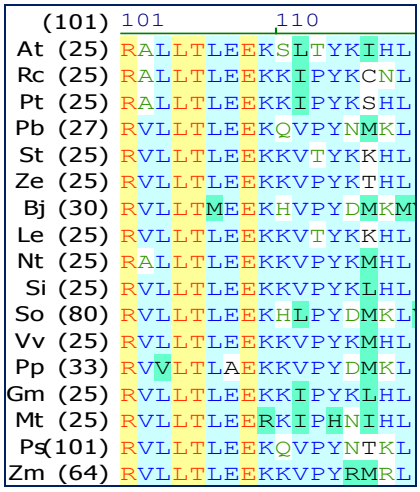


Figure 3.2: Conservation of the N-terminal domain ($RAx_{13}HL$) of DHAR1
Sequences of plant DHAR1 protein homologs, identified by BLAST and aligned using AlignX (Vector NTI, Invitrogen, color background: yellow, identical aa; blue, conservative aa; white, weakly similar aa; green, block of similar aa. The species abbreviations are as follows: At, *Arabidopsis thaliana*; Bj, *Brassica juncea*; Gm: *Glycine max*; Le, *Lycopersicon esculentum*; Mt, *Medicago truncatula*; Nt, *Nicotiana tabacum*; Pp, *Physcomitrella patens* subsp. patens; Pb, *Pinus bungeana*; Ps, *Pisum sativum*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Si, *Sesamum indicum*; So, *Solanum lycopersicum*; St, *Solanum tuberosum*; Tp, Vc, *Volvox carteri* f. nagariensis; Vv, *Vitis vinifera*; Ze, *Zinnia elegans*; Zm, *Zea*

Several GSTs (GSTU19, GSTU20, GSTF7 and GSTF10) were identified in *Arabidopsis* leaf peroxisomes by proteome analyses [see 1.1.1.2, (Reumann et al., 2009)]. However, the four GSTs lacked any predictable PTSs. To validate peroxisome targeting of GSTs, the full-length proteins (GSTU19, At1g78380; GSTU20, At1g78370; GSTF7, At1g02920; GSTF10, At2g30870) were fused N-terminally with EYFP. All reporter-fused proteins remained in the cytosol upon transient expression in onion epidermal cells (Figure 3.1, D-H, 18-48 h expression time). GSTF10 was detected also in organelle-like structures in a few cells (Figure 3.1, G, 18-48 h expression time). However, the identity of these subcellular structures could not be investigated because of the low efficiency of organelle targeting. To better investigate GSTF10, EYFP-GSTF10 was transiently expressed in tobacco protoplasts and appeared to remain in the cytosol (Figure 3.1, K, 24-48 h).

Finally, another detoxification enzyme [S-hydroxymethyl glutathione dehydrogenase/S-nitrosoglutathione reductase (HMGDH/GSNOR, At5g43940)] was also detected in *Arabidopsis* leaf peroxisomes by proteome analyses (Reumann et al., 2007). HMGDH is important in controlling S-nitrosoglutathione turnover, and was reported to afford pathogen resistance in *Arabidopsis* (Rusterucci et al., 2007). The full-length HMGDH was fused N-terminally with EYFP and remained in the cytosol upon transient expression in onion epidermal cells and tobacco protoplast (Figure 3.1, I and L, 18-48 h).

3.2.2 Isolation of homozygous *gr1* and *dhar1* mutants

To initiate physiological functional studies for peroxisomal proteins (GR1 and DHAR1, see 1.1.1.2), homozygous mutants were isolated from *Arabidopsis* T-DNA insertion lines for GR1 and DHAR1 (see 2.2.1.5). T-DNA insertion mutants (see 2.2.1.5) were screened and identified using T-DNA Express (http://signal.salk.edu/cgi-bin/tdnaexpress). T-DNA insertions were generated in the wt Col-0 background. The T-DNA was located at the 3rd of 15 introns in *gr1* and 3rd of 3 exons in *dhar1*. In order to obtain homozygous plants of the mentioned T-DNA insertion lines, a number of genomic PCRs were carried out using genotyping primers which were designed using T-DNA Primer Design tool (http://signal.salk.edu/tdnaprimers.2.html). Genomic DNA was subjected to PCR using the two gene-specific primers (LP and RP) together with the T-DNA specific primer (LBa1: SALK). Several homozygous mutant plants were identified for *gr1* and *dhar1* (Figure 3.3, A).

3.2.3 Analysis of metabolic peroxisome functions in *gr1* and *dhar1* mutants

Photorespiration is accomplished by chloroplasts, peroxisomes, and mitochondria (see 1.1.1.1). Mutants that have a stronger growth defect phenotype in normal air (360 ppm CO_2) are usually characterized as photorespiration mutant if the phenotype is less obvious in high CO_2 conditions, e.g., 670 ppm [e.g. *pex14* null mutant (Orth et al., 2007; Zhang and Hu, 2009)], because photorespiration is not required under high CO_2 conditions (Reumann and Weber, 2006; Kaur et al., 2009). Because *gr1* and *dhar1* showed no growth defect phenotype in normal air, they were investigated for their photorespiration activity by incubating different plants (wt Col-0, *gr1*, and *dhar1* plants) in both low CO_2 concentration (80 ppm) and ambient air (see 2.2.5.3). *gr1*, and *dhar1* plants grew similar to the wt Col-0 under both conditions (Figure 3.3, B). These data however experimentally done once, but indicate that GR1 and DHAR1 don't have any indirect impact on photorespiration.

To determine whether the disruption of *GR1* and *DHAR1* negatively affected seedling establishment, hypocotyls lengths of dark-grown seedlings (wt Col-0, *gr1*, and *dhar1*) were measured upon seed

germination in the presence or absence of sucrose (see 2.2.5.1). The *pex14* null mutant, which is defective in PEX14 (see 1.1.2) which is involved in peroxisomal matrix protein import (see Figure 1.5), has a sugar-dependent phenotype (Orth et al., 2007), was used as a control. On sucrose-free medium, hypocotyl elongation was slightly inhibited in wt Col-0, *gr1* and *dhar1* mutants as compared to sucrose-containing media (Figure 3.3, C). In contrast, hypocotyl elongation was largely inhibited for *pex14* seedling (Figure 3.3, C). These data indicate that both GR1 and DHAR1 are not involved in lipid β-oxidation (see 1.1.1.1).

Next, *gr1* and *dhar1* seedlings were treated by IBA (see 2.2.5.2) to further dissect any possible defect in β-oxidation. IBA is a protoauxin that can be metabolized to the bioactive auxin IAA through peroxisomal β-oxidation in wt Col-0. Mutants deficient in β-oxidation are resistant to the inhibitory effect of IAA on primary root elongation (Hayashi et al., 1998; Zolman et al., 2001). High levels of IBA inhibited root elongation in *gr1* and *dhar1* seedlings, and showed no significant resistance to the auxin, compared with the wt Col-0 plants (Figure 3.3, D). The *pex14* mutant, which was used as a positive control, was resistant to the inhibition of root elongation by IBA (Figure 3.3, D) over a range of concentrations (5-10 µM), consistent with previous reports (Orth et al., 2007; Zhang and Hu, 2010). These data indicate that both GR1 and DHAR1 are not involved in IBA-to-IAA metabolism.

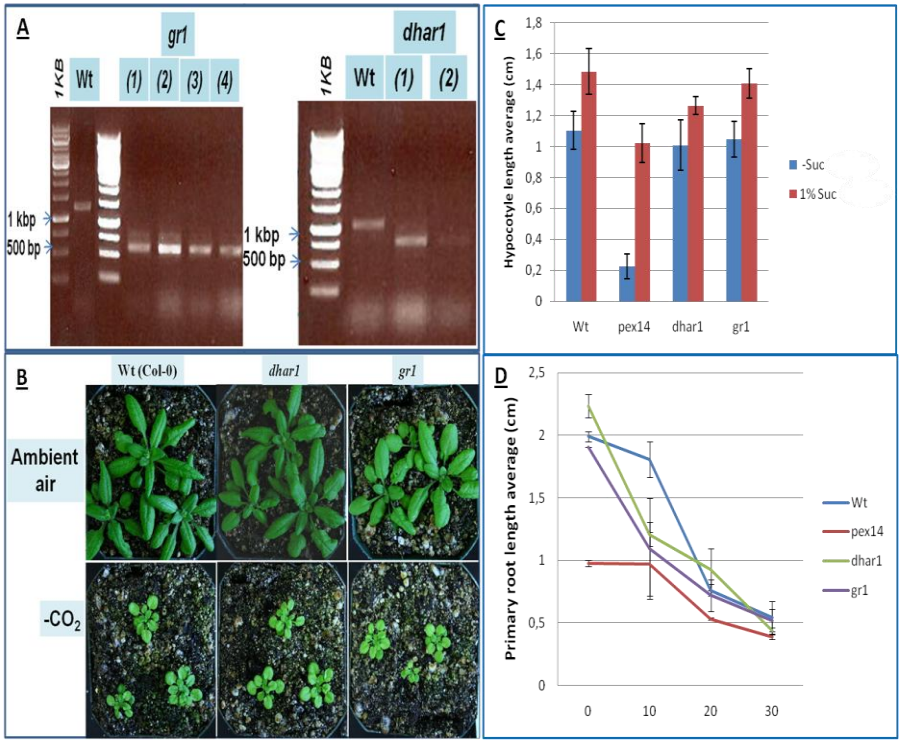


Figure 3.3: Metabolic assays applied to *gr1* and *dhar1* mutants

A: Identification and characterization of homozygous mutants for *gr1* and *dhar1* by genomic PCR. Specific primers were used for genotyping (LP+LBa1+RP). In the lanes of *gr1* (1-4) the T-DNA-specific band size of ~600 bp was present, and the wt band of the size 1200 bp was absent. In the lanes of *dhar1* (1-2) the T-DNA-specific band size of ~700 bp was present, and the wt band of the size 1100 bp was absent. B: Photorespiration assay (n=1) where wt (Col-0) and mutants were planted on MS agar plates, and after 2 weeks were transferred to soil in duplicates (one to grow in ambient air, and the second to grow in low CO_2, 80 ppm). C: Sucrose dependence assay (n=3). Seedlings were grown on half-strength LS with vitamins (with or without 1% (w/v) sucrose) for 6 d in the dark, then the length of 10-15 hypocotyls was measured using the ImageJ program (see 2.2.5.1). Average values of hypocotyl lengthes were calculated for each mutant and are shown. D: Effect of IBA on primary root elongation (n=2). Plants were grown for 7 d in the light on half-strength LS media supplemented with 0, 10, 20 and 30 µM IBA (X-axis). The length of 10-15 primary roots was measured using imageJ program and averaged (see 2.2.5.2). B was done once, while C and D were repeated 3 and 2 times with similar results, respectively. Bars represent SD, for calculations see 2.2.5, for each assay.

3.2.4 Production of recombinant proteins for GR1 and DHAR1

To be able to study the physiological function and the kinetic parameters of *Arabidopsis* GR1, DHAR1 and GSTT1, the tagged recombinant proteins [MBP and His_6] were produced for affinity purification. To this end, the full-length cDNAs of GR1, DHAR1 and GSTT1 were subcloned in two different vectors, pMALc2X and pQE31 (see 2.1.3), to generate N-terminally tagged fusion proteins, with MBP or His_6 tags, respectively. The constructs in pMALc2X were transformed and expressed in *E. coli* BL21 (see 2.1.2) cells (30-37^{o}C mid-log grown cultures). The recombinant proteins were detected from the IPTG induced cultures using 12% SDS-PAGE. The theoretical recombinant protein sizes were calculated (MBP-GR1, 96.97 kDa; MBP-DHAR1, 66.4 kDa; MBP-GSTT1, 70.75 kDa), and all recombinant proteins were successfully detected based on their sizes (Figure 3.4, A and C). The pQE31-based constructs were transformed and expressed in SG13009 *E. coli* cells (30-37^{o}C mid-log grown cultures, (see 2.1.2)). The recombinant proteins for DHAR1 and GSTT1 were successfully expressed and produced His_6-DHAR1 (24.4 kDa) and His_6-GSTT1 (28.75 kDa) proteins (Figure 3.4, B and C), while His_6-GR1 (54.97 kDa) was not detected (Figure 3.4, B).

To determine the solubility of the recombinant tagged proteins, a single colony of the *E. coli* cells carrying each of the recombinant plasmids was grown in LB medium and induced. After sonication, the bacterial lysates were centrifuged to subfractionate the cells into an insoluble and a soluble fraction. After resuspension of the pellets, both fractions were subjected to the same treatment and the recombinant proteins were run on 12% SDS-PAGE (Figure 3.4, C). Five recombinant proteins (i.e., 5) were found to be partially soluble. The availability of GR1 and DHAR1 recombinant proteins with two different tags will allow further studies in order to identify their physiological roles *in vitro*. The next step is to purify the recombinant proteins, and possibly cleave the MBP tag by factor protease Xa (see 2.1.3) and to investigate the kinetic characteristic for the selected proteins. Additionally, the entire ASC-GSH cycle activity shall be investigated in *Arabidopsis* leaf peroxisomes that were isolated in the course of this study (see 2.2.7).

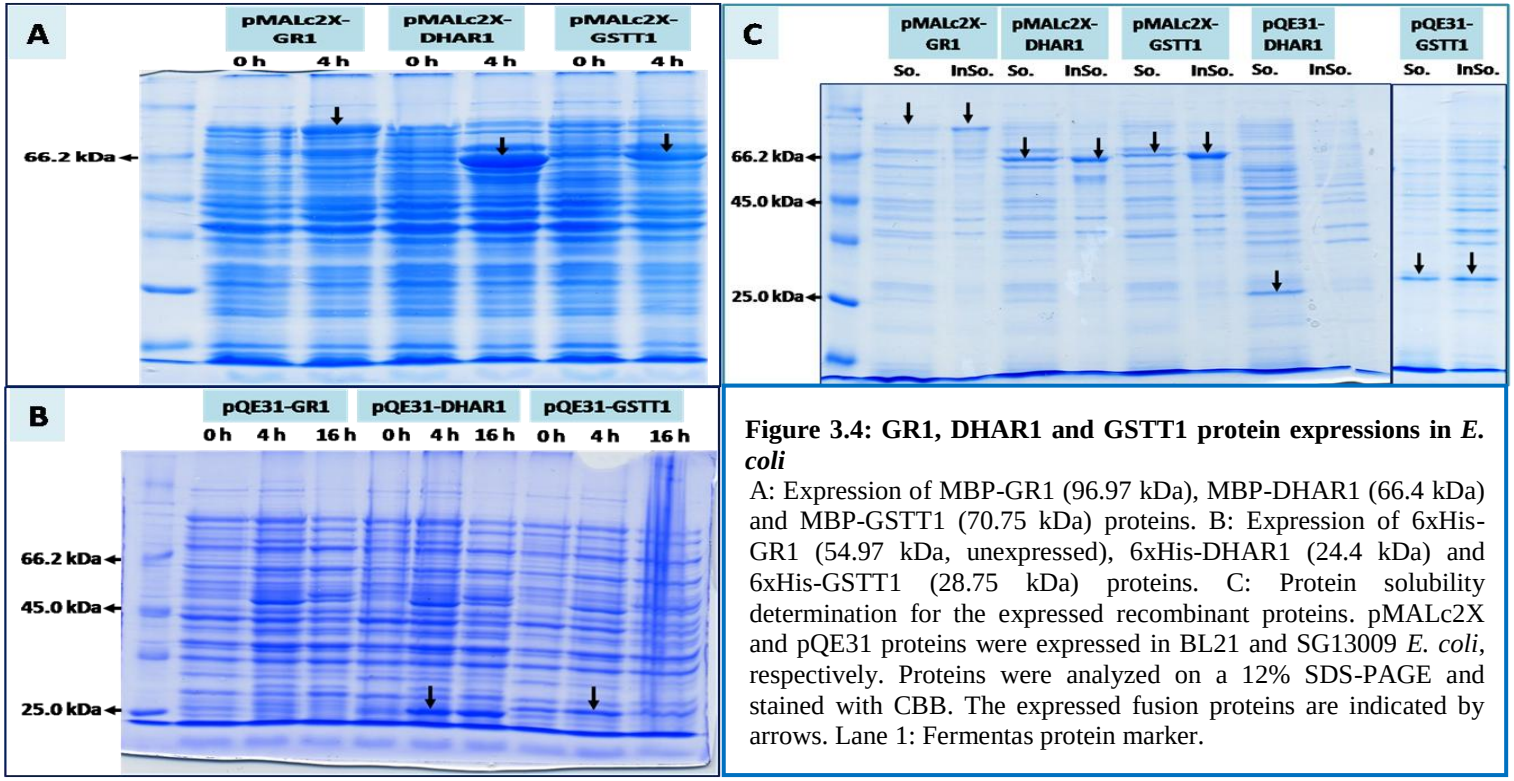


Figure 3.4: GR1, DHAR1 and GSTT1 protein expressions in *E. coli*

A: Expression of MBP-GR1 (96.97 kDa), MBP-DHAR1 (66.4 kDa) and MBP-GSTT1 (70.75 kDa) proteins. B: Expression of 6xHis-GR1 (54.97 kDa, unexpressed), 6xHis-DHAR1 (24.4 kDa) and 6xHis-GSTT1 (28.75 kDa) proteins. C: Protein solubility determination for the expressed recombinant proteins. pMALc2X and pQE31 proteins were expressed in BL21 and SG13009 *E. coli*, respectively. Proteins were analyzed on a 12% SDS-PAGE and stained with CBB. The expressed fusion proteins are indicated by arrows. Lane 1: Fermentas protein marker.

3. 3. Identification of defense-related peroxisomal proteins

Recently, peroxsiomes were reported to have roles in innate immunity and plant resistance against pathogens (see 1.1.1.3). To be able to address the mode of action of peroxisomes in plant defense mechanisms, it is important to characterize additional possible peroxisomal defense-related proteins. Interestingly, several defense-related proteins were predicted to contain putative PTS1s after the application of PWM and RI models to the *Arabidopsis* genes [(Lingner et al., 2011), see Table 3.2]. The proteins of interest were selected based on their annotation as defense-related in plants, human, and bacteria (Table 3.2) and on their probability of carrying predicted PTS1s.

3.3.1 Validation of AtMIF1 targeting to peroxisomes

Macrophage migration inhibitory factor (MIF) is an immune-regulatory protein, and is implicated in several inflammatory diseases in human (Golubkov et al., 2006). Importantly, MIF counter-regulates the immunosuppressive effects of steroids and hence is critical in human immune system both locally and systemically (Golubkov et al., 2006). One of three *Arabidopsis* MIF homologs, in this study referred to as AtMIF1 (SKL>; At3g51660), was identified in *Arabidopsis* leaf peroxisomes by proteome analyses (Reumann et al., 2007), and also was predicted by the PTS1 prediction algorithms (Lingner et al., 2011). Based on the available results of microarray experiments, which are provided by Genevestigator and the eFP browser (www.genevestigator.com; http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi), the expression pattern of *AtMIF1* was investigated. Anatomically, *AtMIF1* transcripts appeared to be restricted to adult and senescent leaves. Developmentally, *AtMIF1* transcripts were restricted to developed rosette and flowers (Figure 3.7, A and B). Several biotic stresses induced *AtMIF1*, for instance, bacteria (virulent and avirulent *Ps*), fungi (necrotrophic: *Botrytis cinerea*), and viruses. Moreover, *AtMIF1* was also induced upon treatment by hormones (e.g., SA, JA, and ABA), and bacterial elicitors (e.g., flg22, LPS, and HrpZ). Based on eFB browser microarray experiments, *AtMIF1* appeared to be induced by an oomycete derived elicitor (GST-NPP), but not by the oomycete (*Phytophthora infestans*) itself, which might indicate that the

pathogen evolved a mechanism to suppress *AtMIF1* induction (Table 3.1). Finally, *AtMIF1* also appeared to be expressed by abiotic stresses (cold, drought, osmosis, and wounding, Figure 3.7, C). The microarray-based expression analyses support the prediction of AtMIF1 as an important protein in *Arabidopsis* defense responses.

In order to verify the presence of AtMIF1 in peroxisomes, full-length AtMIF1 was fused N-terminally with EYFP. Upon transient expressions in both onion epidermal cells and tobacco protoplasts, the fusion protein was targeted to peroxisomes (Figure 3.5, A and B). Additionally, two *Arabidopsis* homologs of AtMIF1 were identified by Blast search for AtMIF1 paralogs (AtMIF2: ATL>, At5g01650.1 and AtMIF3: STF>, At5g57170). Both AtMIF1 homologs were detected in chloroplasts by proteome analysis (Zybailov et al., 2008). In contrast to *AtMIF1*, both *AtMIF2* and *AtMIF3* appeared to be more constitutivly expressed, and were very slightly induced by light stress and biotic stresses (Genevestigator, data not shown). AtMIF2 has a PTS1-like tripeptide (ATL>); PWM score 0.48 [updated according to (Lingner et al., 2011)] which is close to the PTS1 prediction threshold (0.412). To be able to address if AtMIF2 is also targeted to peroxisomes, the full-length cDNA was subcloned in the back of EYFP. However, EYFP-MIF2 remained in the cytosol upon transient expression in onion epidermal cells (data not shown).

3.3.2 Validation of AtSurE targeting to peroxisomes

The stationary phase survival protein (SurE) has activities as nucleotidase and exopolyphosphatase and is thought to be involved in stress responses in *E.coli* (Proudfoot et al., 2004). One *Arabidopsis* SurE homolog, here referred to as AtSurE (SSL>; At4g14930) was predicted by the PTS1 prediction algorithms (Lingner et al., 2011). Investigation of *Arabidopsis* microarray experiments (by eFP and Genevestigator) showing expressions of *AtSurE*, indicated that *AtSurE* is constitutively expressed, and is highly induced in response to biotic (bacteria: e.g., *Pst* DC3000 and viruses) and abiotic stresses (drought, heat, osmosis and salinity, Figure 3.7 and Table 3.1). Thus, AtSurE could be related to *Arabidopsis* stress responses.

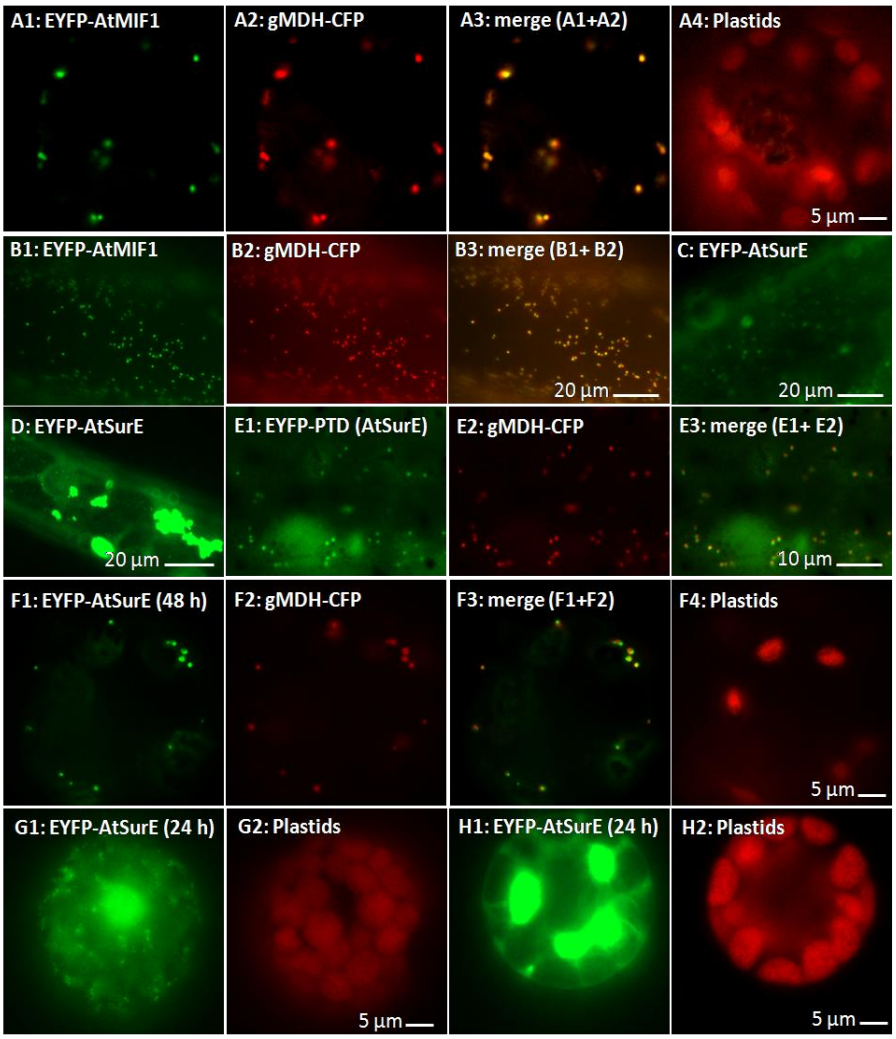


Figure 3.5: *In vivo* subcellular localization of AtMIF1 and AtSurE
The full-length proteins of AtMIF1 and AtSurE were fused N-terminally with EYFP. The fusion genes were then transiently expressed in onion epidermal cells and tobacco protoplasts. AtMIF1 was detected in peroxisomes in both expression systems (A and B). AtSurE was detected in peroxisomes 48 h P.T. in protoplasts (F), while the fusion protein was detected in unknown organelle-like structures after 24 h and in onions (C and G) and in aggregates (D and H). EYFP-PTD of AtSurE was also detected in peroxisomes in onions (E). Peroxisomes were labeled with gMDH-CFP (Fulda et al., 2002). The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown, except for A (n=2). Expression times (18-48 h).

The full-length AtSurE cDNA was fused in the back of EYFP, and the fusion protein was targeted weakly (mostly at the detection limit) to organelle-like structures upon transient expression in onion epidermal cells (Figure 3.5, C, 18-48 h expression time). Sometimes, the yellow fluorescent organelles were found aggregated in distinct locations within the cells, which remained alive (Figure 3.5, D, 18-48 h expression time). Often, the aggregates were very large with a diameter of 20-40 µM, indicating that a large number of small punctate structures must have aggregated together, or that the fusion protein accumulated somehow intensively in these structures and failed to be exported. The identity of these aggregate-like structures remains elusive, because of the absence of convincing coincidence with CFP-labeled peroxisomes.

Upon transient expression in tobacco protoplasts the same fusion protein was found to change its subcellular localization in a time-dependent manner. Twenty-four h post transformation (P.T.) and similar to onions, the fusion protein was detected in organelle-like structures of smaller size as compared to standard leaf peroxisomes in tobacco protoplasts. Some large yellow fluorescent clusters were observed of a size of ca. 20-40 µM (Figure 3.5, G and H). The coincidence of EYFP-labeled structures with the CFP-labeled peroxisomes could not be approved at 24 h P.T. in protoplasts (Figure 3.5, G). However, the yellow fluorescent organelles of tobacco protoplasts reproducibly coincided with CFP-labeled peroxisomes in a low but significant number of cells 48 h P.T. (Figure 3.5, F). Taken together, the data indicate that AtSurE was targeted first and primarily to unknown structures and subsequently to peroxiomes by an unknown mechanism. These data prompted us to address if AtSurE is indeed targeted to peroxisomes by the predicted PTS1 by constructing EYFP-PTD (SSL>). As predicted, the domain construct was targeted to organelle-like structures upon transient expression in onion epidermal cells, nearly all of which coincided with CFP-labeled peroxisomes (Figure 3.5, E). However, a significant cytosolic background staining of EYFP was noticed.

3.3.3 Validation of additional defense-related proteins

Another five *Arabidopsis* defense-related candidate proteins (Table 3.2) were predicted by PTS1 protein prediction models (Lingner et al., 2011). Two identified candidates belong to NBS-LRR R proteins (see 1.2.1.2 and Figure 1.8). One protein is the *Arabidopsis* LIM domain-containing protein (here referred to as: AtLIMDP, variant 1) is encoding 1613 aa and was named recently as chiling sensitive 3 (CHS3)/DA1-related protein 4 (CHS3/DAR4) and contains a TIR-NBS-LRR domain at the N terminus, and two LIM domains at the C-terminus (Yang et al., 2010). The second protein, which has a CC-NBS-LRR domain, is the *Arabidopsis* disease resistance protein (referred to as AtDRP, variant 2). AtDRP had not yet been investigated and was annotated to be involved in defense response based on its domain structure similarities to R proteins (Meyers et al., 2003). *Arabidopsis* Cinnamyl-alcohol dehydrogenase 7 (here referred to as AtCAD7, varaiant 2), was also called Elicitor-activated gene 3-1 (ELI3-1) protein that was originally identified as part of the defense response in parsley after treatment by fungal elicitor (Somssich et al., 1989). In another study, ELI3-1 was also expressed and isolated from *Arabidopsis* treated by fungal elicitor (Trezzini et al., 1993). AtCDR1 was reported to be functional as a highly specific aspartic proteinase (Simöes et al., 2007). Moreover, AtCDR1 is involved in signaling of disease resistance (see SAR, 1.2.2.3, (Xia et al., 2004)). In addition to these defense candidates identified by PTS1 prediction, *Arabidopsis* ozone induced protein 1 (AtOZI1) was identified in *Arabidopsis* leaf peroxisomes by proteome analyses (Reumann et al., 2007) and is lacking any predictable targeting signals. Overall, microarray experiments of the identified five defense candidates indicated that all of them were constitutivly expressed except for *AtCDR1*, but also all appeared to be induced upon different biotic and abiotic stresses (Figure 3.7 and Table 3.1).

The full-length cDNAs of AtCAD7, AtDRP, and AtCDR1 were fused in the back of EYFP. Regarding the long protein AtLIMDP, its C-terminal domain comprising amino acid 1141 to 1613 (472 aa) was fused in the back of EYFP to facilitate the subcloning. To this end, this region was amplified from the full-length RIKEN cDNA (see 2.2.2.1) by PCR. The cDNA, however, contained an additional (T) nucleotide at

position 4561 and introduced a frame shift in the final reporter gene construct.

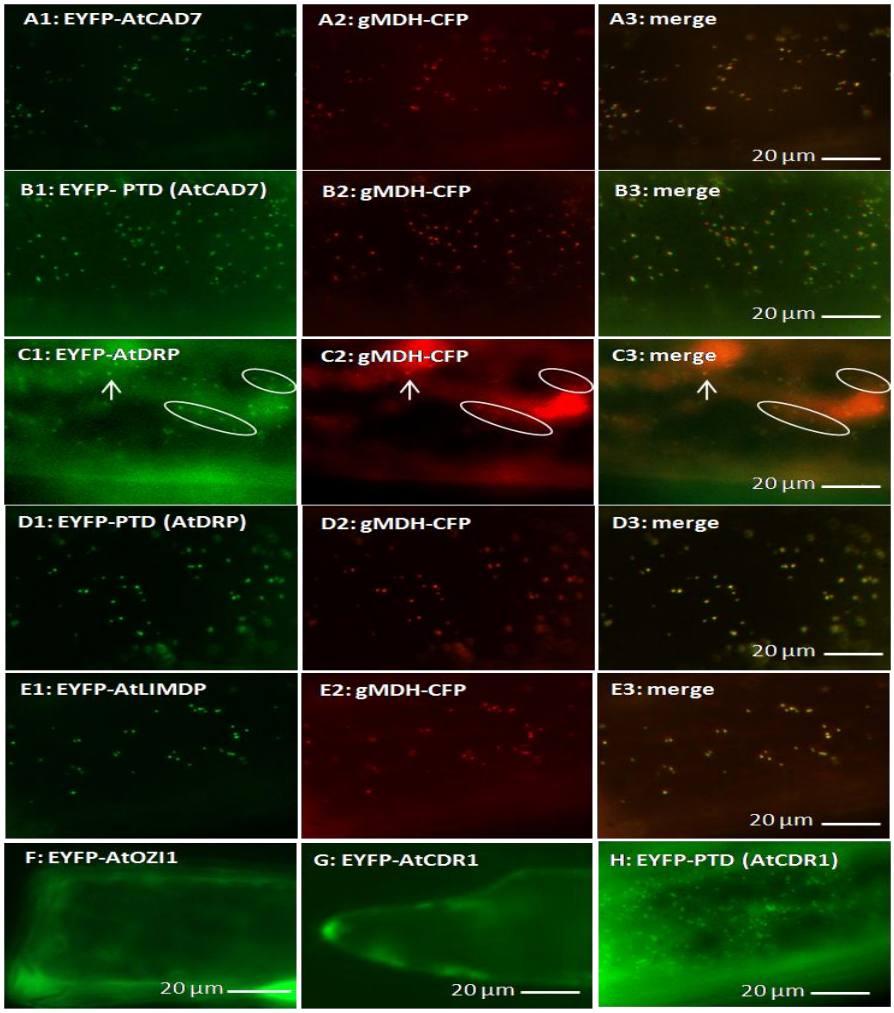


Figure 3.6: *In vivo* subcellular targeting of putative defense proteins
The full-length proteins (AtCAD7, AtDRP, AtLIMDP, AtOZI1, and AtCDR1) were fused N-terminally with EYFP. The fusion proteins were transiently expressed in onion epidermal cells. AtCAD7, AtDRP and AtLIMDP were detected in peroxisomes (A, C, and E), while AtOZI1 and AtCDR1 remained in the cytosol (F and G). Moreover, the EYFP-PTDs of AtCAD7, AtDRP and AtCDR1 were detected in peroxisomes upon expression in onions (for AtCDR1, only single labeling data are available, H). Peroxisomes were labeled with gMDH-CFP (Fulda et al., 2002). The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown, except for C (n=2). Expression times are 18 h for A; B; D; E, and 1 week for C; G-H.

The additional nucleotide was removed from the *EYFP-LIMDP* by SDM of the full vector (see 2.2.2.7). Moreover, AtOZI1 full-length cDNA subcloning in the back of EYFP and preliminary fluorescence microscopy was done by a bachelor student (Amundsen, 2009). Three reporter fusions (AtCAD7, AtDRP, and AtLIMDP) were targeted to organelle-like structures upon transient expression in onion epidermal cells, and the organelles coincided with CFP-labeled peroxisomes (Figure 3.6, A, C and E, respectively). By contrast, the AtOZI1 and AtCDR1 fusion proteins remained in the cytosol (Figure 3.6, F and G, 18 h-1 week expression time). In addition to the full-length protein targeting, confirmation of the predicted PTS1 tripeptides was accomplished by constructing three EYFP-PTD fusions (AtCAD7, SHL>; AtDRP, CRL> and AtCDR1, AKM>). As predicted all three domain constructs were targeted to organelle-like structures that coincided with CFP-labeled peroxisomes (Figure 3.6, B, D and H), demonstrating that all three proteins carry functional PTS1 domains. Efficient peroxisome targeting of the EYFP-PTD of AtDRP in particular supported peroxisome targeting of the expressed full-length fusion protein in onion epidermal cells.

Table 3.1: Gene expression analyses for defense-related genes
The expression data derive from microarray experiments and were retrieved using the eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi, BAR, Toronto). Expression symbols represent the expression pattern for the genes of interest upon biotic stress treatments (0 for uninduced and (+, ++) for induced). The symbols were based on the "electronic-fluorescent pictographic" representations of gene expression patterns (Schmid et al., 2005).

Treatment/ Gene	*AtMIF1*	*AtSurE*	*AtCDR1*	*AtLIMDP*	*AtDRP*	*AtCAD7*	*AtOZI1*
Botrytis cinerea	+	0	0	0	++	0	++
Phytophthora infestans	0	0	+	+	0	0	++
Erysiphe orontii	0	0	0	0	0	0	++
PST DC3000	++	++	0	0	0	++	+
PST DC3000 (avrRpm1)	++	+	0	0	0	+	+
PST DC3000 (hrcC-)	+	0	0	++	0	0	+
Ps Phaseolicola	++	0	0	++	0	0	+
flg22	++	0	+	0	0	0	+
HrpZ	++	0	0	+	+	0	+
LPS	+	0	0	0	0	0	0
GST-NPP1	++	0	0	++	++	0	++
SA	+	0	0	+	+	+	0
ABA	++	++	0	0	+	+	0
MJ	++	0	+	0	0	0	0

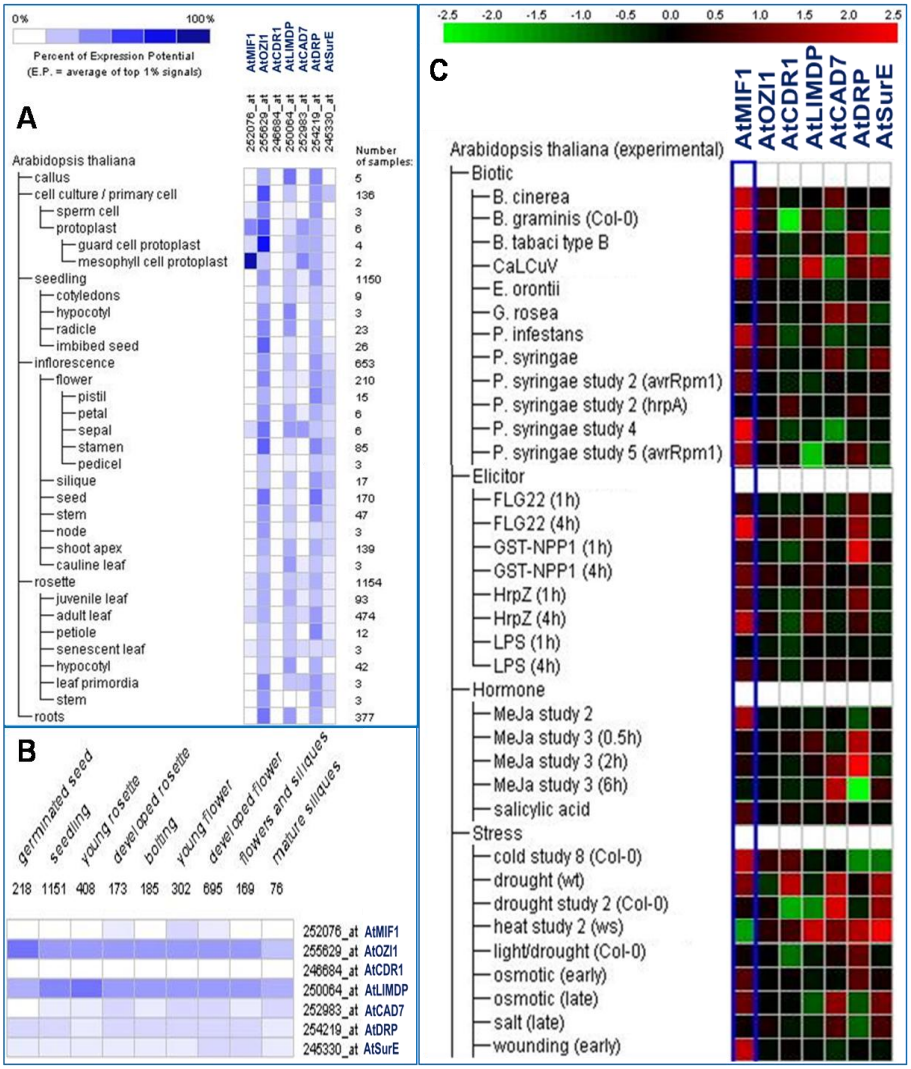


Figure 3.7: Gene expression analyses of defense-related genes
Gene expression analyses of the seven *Arabidopsis* defense-related genes, which were investigated in the present study. A, B and C are images representing anatomy, development, and stress-related expressions, respectively. The expression data derived from microarray experiments and were retrieved using Genevestigator (www.genevestigator.com; (Zimmermann et al., 2004)). High and low expression levels are reflected semi-quantitatively by *dark* and *light* coloring, respectively.

Table 3.2: Summary of subcellular localization data for defense proteins

AGI code	Acronym	Subcellular localization		Annotation	Data source	PTS1
		Onions	**Tobacco protoplasts**			
At3g51660	AtMIF1	Peroxisomes	Peroxisomes	Macrophage migration inhibitor factor homolog	Reumann et al. (2007)	SKL>
At4g14930	AtSurE	Unknown organelles	Peroxisomes	Acid phosphatase survival protein SurE	PTS1 prediction	SSL>
At4g14930	PTD (AtSurE)	Peroxisomes	n.d.			
At5g33340	AtCDR1	Cytosol	n.d.	Consititutive disease resistance 1; aspartic-type endopeptidase/pepsinA	PTS1 prediction	AKM>
At5g33340	PTD (AtCDR1)	Peroxisomes	n.d.			
At5g17890.1	AtLIMDP	Peroxisomes	n.d.	LIM domain-containing protein / chiling sensetive 3 (CH3)/DA1-related protein 4 (DAR4)	PTS1 prediction	SKL>
AT1G58807.2	AtDRP	Peroxisomes	n.d.	Disease resistance protein-related	PTS1 prediction	CRL>
At1g58807.2	PTD (AtDRP)	Peroxisomes	n.d.			
At4g37980.2	AtCAD7	Peroxisomes	n.d.	Cinnamyl-alchol dehydrogenase7/Elicitor-activated gene (ELI3-1)	PTS1 prediction	SHL>
At4g37980.2	PTD (AtCAD7)	Peroxisomes	n.d.			
At4g00860	AtOZI1	Cytosol	n.d.	Ozone-induced protein	Reumann et al. (2007)	??

3. 4. NHL protein family investigations

Initially, 28 *Arabidopsis* NHL members (see 1.2.3.1) were identified (Dormann et al., 2000). Upon the completion of the *Arabidopsis* genome sequencing (Arabidopsis genome initiative, 2000), NHL family members were found to be 45 genes including *NDR1* (Zheng et al., 2004). Three proteins, NHL4, NHL6 and NHL25 were found to carry predicted PTS1 tripeptides (Table 3.3) according to the newly developed PWM and RIM prediction methods [see 3.1.1, (Lingner et al., 2011)]. The three NHL homologs are located in one phylogenetic clade (Figure 3.8, (Dormann et al., 2000; Zheng et al., 2004)). Four additional NHL family members (here referred to as NHL39, NHL39H1, NHL13H1, and NHLx) were identified by bioinformatics domain analysis and noticed to carry possible PTS1 tripeptides, predicted by lower prediction scores (Reumann, unpubl. data, Table 3.3).

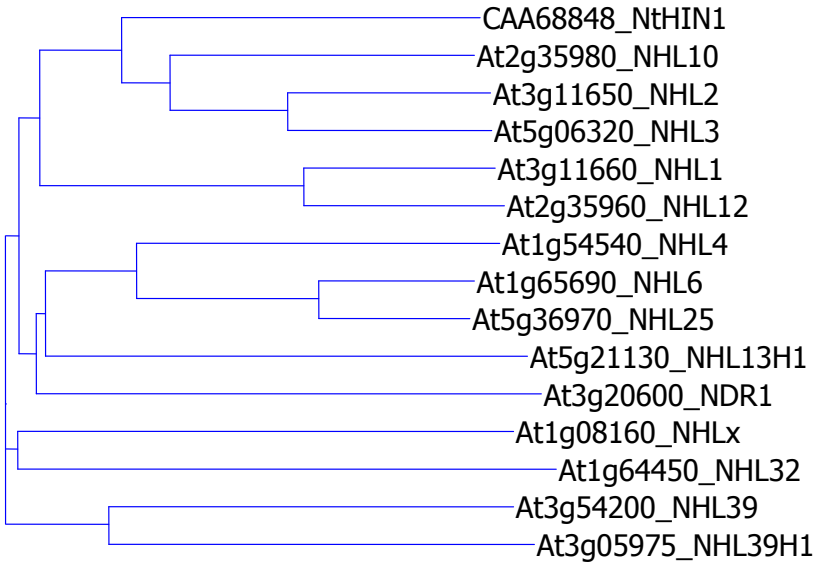


Figure 3.8: Phylogenetic relationship of selected NHL proteins.
To investigate phylogenetic relationship (Dormann et al., 2000; Zheng et al., 2004) among NHL proteins carrying predicted PTS1 domains, *Nicotiana tabacum* HIN1 and *Arabidopsis thaliana* NHL homologs were aligned with the predicted NHL proteins. The phylogram was generated by the AlignX program (Vector NTI, Invitrogen) using the Neighbor Joining method (NJ) (Saitou and Nei, 1987). The NJ method works on a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between the sequences.

Table 3.3: PTS1 predictions for NHL homologs.
The threshold of the prediction scores for predicted peroxisome targeting are for PWM=0.412 and for RIM= 0.219 (Lingner et al., 2011).

AGI code	Acronym	C-terminal tripeptide	PWM score	RIM score
At1g54540	NHL4	AKL>	2.67	0.61
At1g65690	NHL6	LRL>	1.91	0.17
At5g36970	NHL25	FRL>	1.99	0.37
At5g21130	NHL13H1	SLL>	1.63	-0.24
At3g54200	NHL39	TKL>	1.48	-0.02
At3g05975	NHL39H1	TKL>	1.76	0.001
At1g08160	NHLx	TRL>	1.51	-0.17

To this end, the deduced protein sequences of the proposed PTS1 NHL proteins (Table 3.3) were aligned with representative NHL members from the different clustered groups which were reported previously (Dormann et al., 2000; Zheng et al., 2004), and the characterized pathogen-related proteins: NHL2, NHL3 and NHL10 (Century et al., 1995; Gopalan et al., 1996; Dormann et al., 2000; Varet et al., 2002), in order to investigate protein characteristics for NHL4, NHL6 and NHL25 proteins. Obviously, the three motifs (see 1.2.3.1 and Figure 3.9) conserved among *Arabidopsis* NHL proteins were also found in NHL4, NHL6, and NHL25 proteins (Figure 3.9). The Water stress and Hypersensitive response (Why) domain, which was previously identified in HIN1 (Ciccarelli and Bork, 2005), was also found to be conserved in NHL4, NHL6 and NHL25 (Figure 3.9). The WHy domain is comprised of ~100 aa with an alteration of hydrophilic and hydrophobic residues and an almost invariable NPN motif at its N-terminus (Ciccarelli and Bork, 2005). In summary, NHL4, NHL6 and NHL25 share the same protein characteristics and are strongly indicated to have similar functions in plant defense responses.

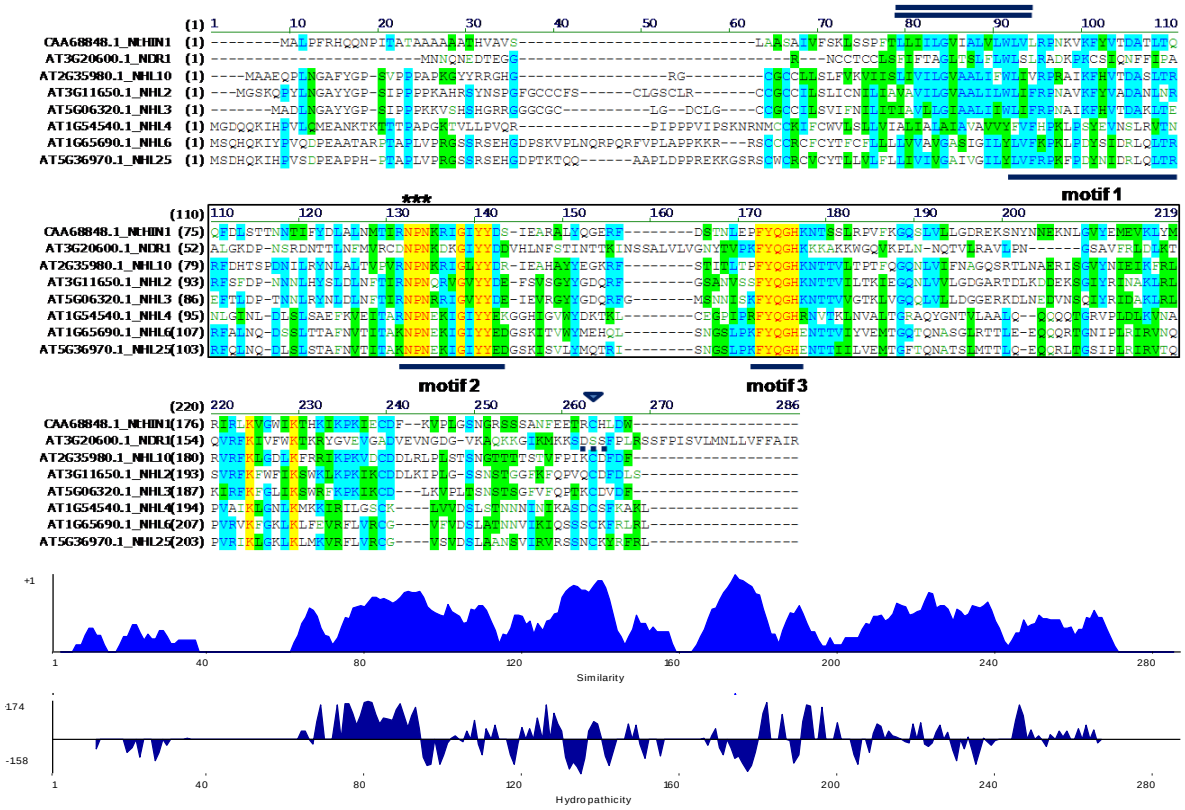


Figure 3.9: Sequence alignment of NHL homologs.
Top panel: Sequence alignment of tobacco HIN1 and selected *Arabidopsis* NHL proteins. The WHy domain (Ciccarelli and Bork, 2005) are boxed and an invariable NPN motif is marked by asterisks. The three conserved sequence motifs among NHL proteins are indicated by bold lines. The hydrophobic anchor sequence is indicated by double lines. An arrowhead refers to the unique GPI anchor of NDR1. Lower panel: the first graph displays the alignment quality profile (similarity). The default values are 1, 0.5 and 0.2 for identical, similar and weakly similar residues, respectively. The second graph displays the hydropathy calculations (Kyte and Doolittle, 1982). Positive numbers indicate hydrophobicity and negative numbers hydrophilicity. The sequence alignment was generated using the AlignX program (Vector NTI, Invitrogen).

3.4.1 *In vivo* subcellular localization of NHL proteins

Based on the PTS1 protein predictions for NHL4, NHL6 and NHL25 proteins (3. 4 and Table 3.3), they were subjected to *in vivo* subcellular localization targeting analyses. The full-length proteins were fused N-terminally with EYFP. *NHL4* was PCR amplified from an available cDNA (ABRC, see 2.2.2.1), while *NHL25* and *NHL6* were amplified by RT-PCR (see 2.2.2.2) from SA-treated *Arabidopsis* leaves and senescent leaves (Figure 3.18), respectively (see 2.2.2.2). The fusion proteins were transiently expressed in onion epidermal cells. Indeed, the three fusion proteins were identified in organelle-like structures. The morphological pattern of these organelles was variable according to their appearance in different transformed cells. The fusion proteins sometimes were very weakly targeted to organelle-like structures, aggregate-like structures, or to both simultaneously (Figure 3.10). However, the detected structures did not coincide with the CFP-labeled peroxisomes (Figure 3.10, A-C, 18 h to 1 week expression times). Moreover, preliminary results confirmed that the organelle-like structures also did not coincide with the CFP-labeled mitochondria in onion epidermal cells, as investigated for NHL4 and NHL25 (data not shown).

Due to its high PTS1 protein prediction score (Table 3.3) and possession of a well-known PTS1 AKL> (Reumann, 2004; Lingner et al., 2011), subcellular targeting of EYFP-NHL4 was also investigated in an alternative expression system, i.e. tobacco leaf protoplasts. The fusion protein was detected in unidentified organelle-like suructures 24 h P.T., but also these organelles mostly did not coincide with CFP-labeled peroxisomes (Figure 3.11, A). But, the fusion protein was clearly identified and coincided with CFP-labeled peroxisomes 48 h P.T. (Figure 3.11, D and E). Astonishingly, different patterns of coincidence of both EYFP-labeled structures with CFP-labeled peroxisomes were detected in different transformed protoplasts: (1) small EYFP-structures appeared to be attached to the surface of CFP-labeled peroxisomes (Figure 3.11, B, C, F, and G), (2) the EYFP fluorescence was detected in small structures that were attached to the surface of CFP-labeled peroxisomes, and faintly in the same CFP-labeled peroxisomes (Figure 3.11, D and H), and (3) EYFP was only

detected in organelles that completely coincided with CFP-labeled peroxisomes (Figure 3.11, E and I).

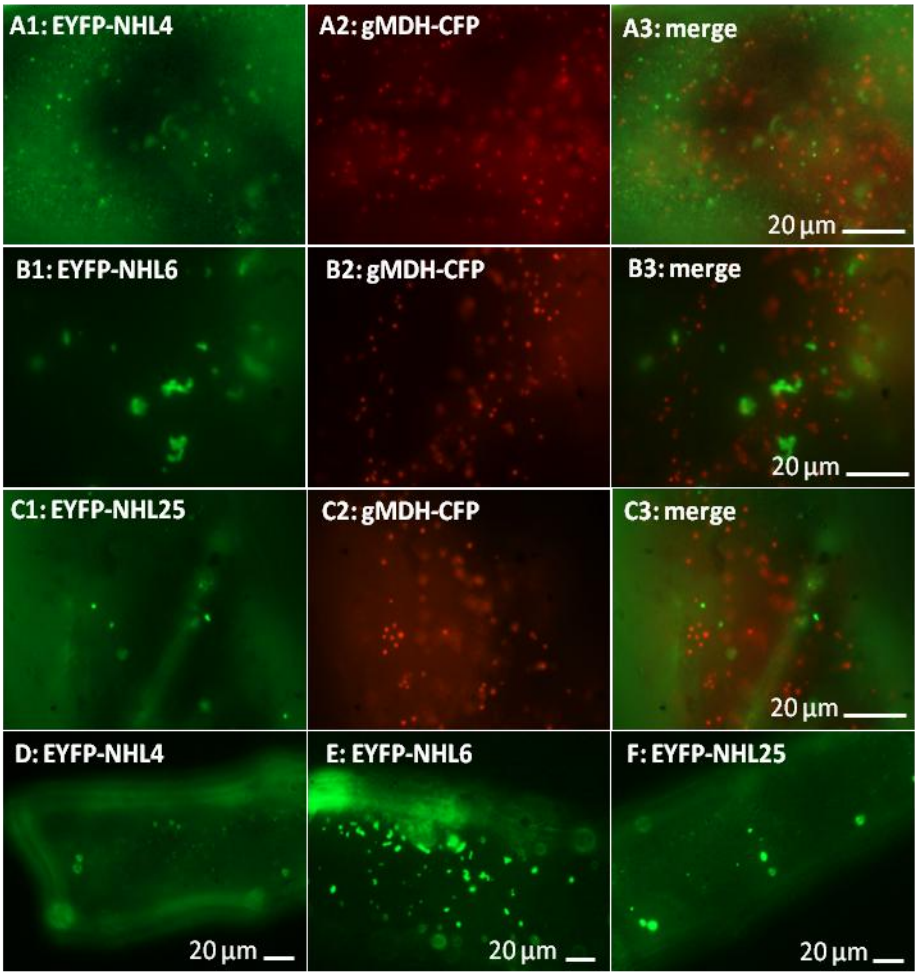


Figure 3.10: *In vivo* subcellular localization of NHL proteins
The full-length proteins of NHL4, NHL6 and NHL25 were fused N-terminally with EYFP. The fusion proteins were targeted to non-peroxisomal unidentified organelle-like structures upon transient expression in onion epidermal cells. In double transformants (A-C), NHL4, NHL6 and NHL25 did not coincide with CFP-labeled peroxisomes. D-F pictures shows formation of aggregate-like structures for NHL4, NHL6 and NHL25. Peroxisomes were labeled with gMDH-CFP (Fulda et al., 2002). The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown, except for B (n=1). Expression times are 18 h for A and 1 week for B-F.

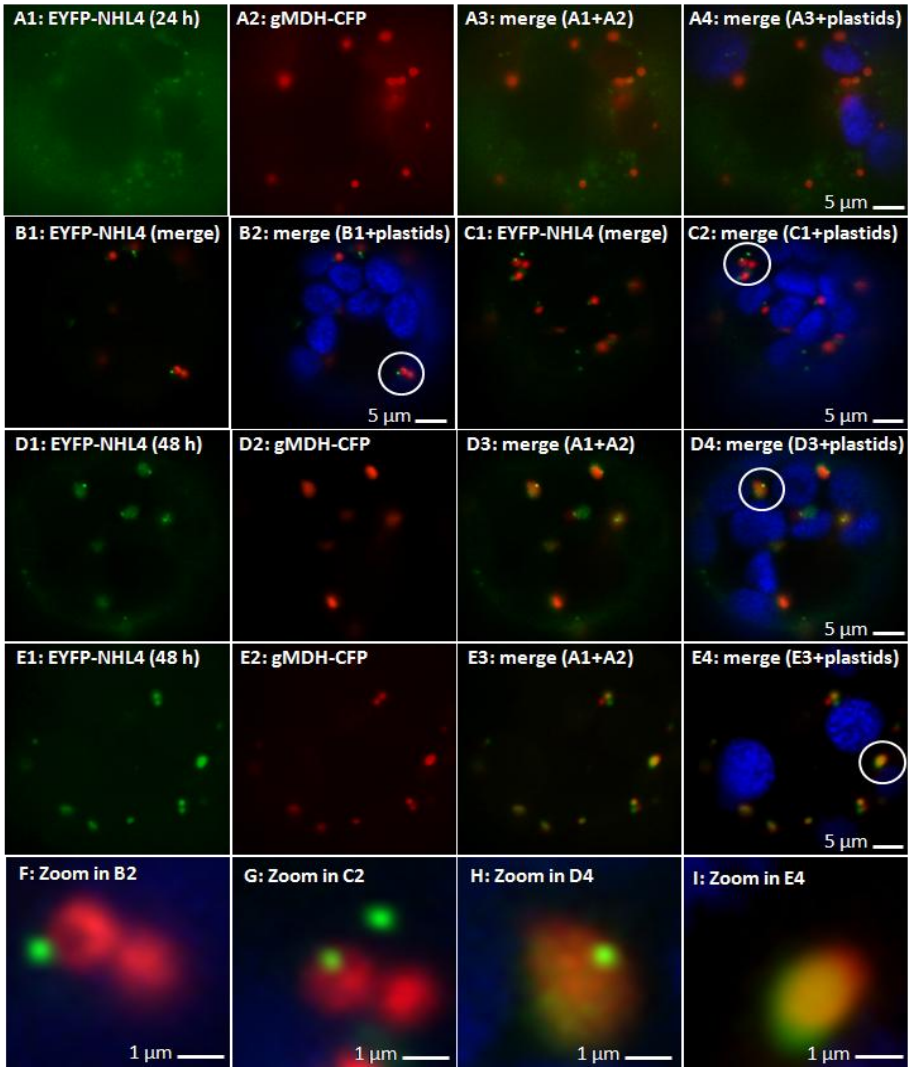


Figure 3.11: *In vivo* subcellular localization of NHL4 in tobacco protoplasts
The full-length NHL4 was fused N-terminally with EYFP. The fusion protein was transiently expressed in tobacco protoplasts. EYFP-NHL4 was detected in organelle-like structures 24 h P.T. (A) and in peroxisomes, 48 h P.T. (B, C, D, and E). B and C are showing surface association of the EYFP-labeled small structures with peroxisomes. D: is detecting EYFP fluorescence in both surface associated peroxisomes and organelle-like small structures. F-I are zoom in/blow-up of the circled single peroxisomes from B2, C2, D4 and E4. Peroxisomes were labeled with gMDH-CFP (Fulda et al., 2002). The cyan fluorescence was converted to red. Plastid autofluorescence was converted to blue. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown. Expression times (24-48 h).

In summary, EYFP-NHL4 was detected in tobacco protoplasts in 1) free small non-peroxisomal organelle-like structures, 2) peroxisome-associated non-peroxisomal organelle-like structures, and 3) peroxisomes (alone or with small EYFP-labeled organelle-like structures attached). Because of the lack of time-lapse imaging for the present study, a targeting mechanism for NHL4 into peroxisomes was hypothesized based on the patterns observed. The EYFP-NHL4 protein could be targeted to peroxisomes in three successive steps, 1) to be targeted to non-peroxisomal unidentified organelle-like structures, 2) the unidentified organelle-like structures associate with the peroxisome surface and 3) the organelle-like structures are releasing their cargo into peroxisomes.

The peroxisomal validation of NHL4 in protoplasts prompted us to construct full-length NHL4 fused N-terminally with CFP. The new CFP fusion protein could be used as a marker to investigate other EYFP-NHL proteins, to determine their coincidence with NHL4 in onion epidermal cells. Indeed, in co-localization experiments, EYFP-NHL25 coincided with the CFP-NHL4 in the same organelle-labeled structures upon transient expression in onion epidermal cells (Figure 3.12, B). It was also indicated (but from preliminary data) that EYFP-NHL6 partially coincided with CFP-NHL4 (Figure 3.12, A). To confirm that NHL4, NHL6 and NHL25 proteins indeed possess functional PTS1 domains as predicted (Table 3.3), the proposed PTDs were used to construct EYFP-PTDs for each of the three proteins. When the EYFP-PTD from NHL6 (LRL>) and NHL25 (FRL>) were transiently expressed in onion epidermal cells, the fusion proteins were targeted to punctate subcellular structures that were validated as peroxisomes by their coincidence with CFP-labeled peroxisomes (Figure 3.12, C and D). Subcloning of the corresponding NHL4 construct remained unsuccessful because of PCR-generated mutations. These data indicate that NHL6 and NHL25 have functional PTS1 domains and PTS1 tripeptides LRL>, and FRL>, respectively. Furthermore, their coincidence with NHL4 in the same subcellular structures in onion epidermal cells indicates that the two NHL proteins are most likely targeted to peroxisomes in tobacco protoplasts, similar to NHL4. This could not be investigated in the present study because of time limitations and needs to be done in the future.

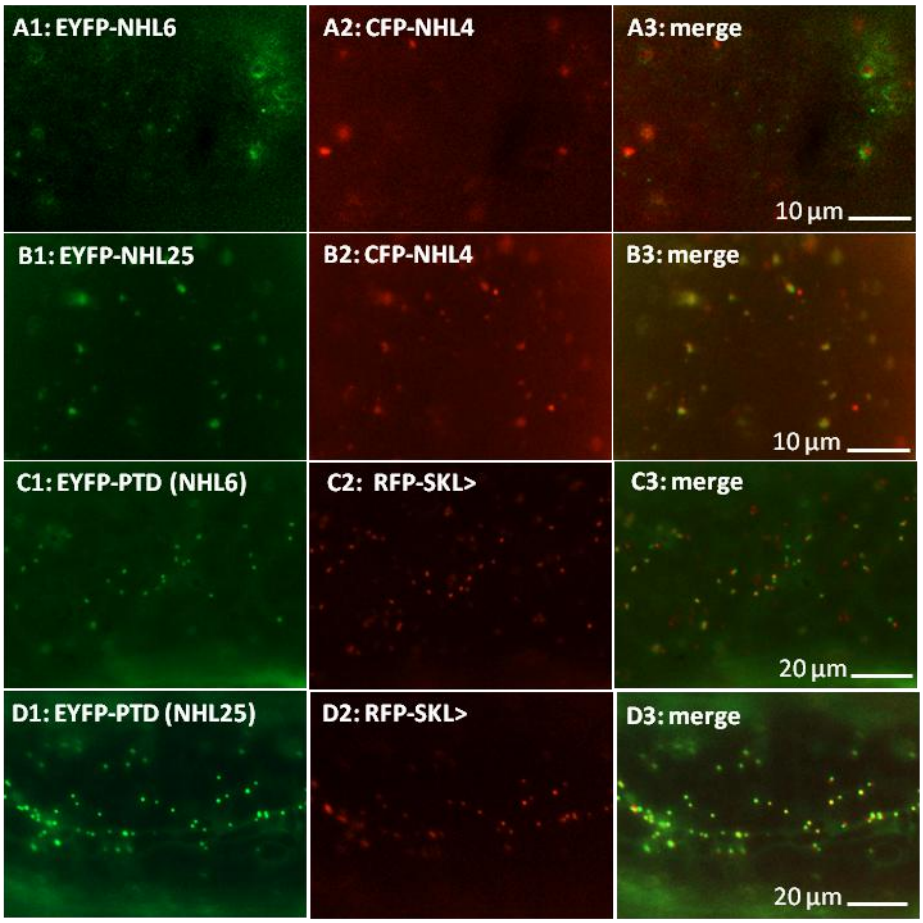


Figure 3.12: *In vivo* subcellular localization of NHL6 and NHL25
A-B: The full-length proteins of NHL6 and NHL25 were fused N-terminally with EYFP, while NHL4 was fused N-terminally with CFP. The EYFP-fusion proteins were co-expressed in onion epidermal cells with CFP-NHL4. Images show partial co-localization for EYFP-NHL6 (A) and complete co-localization of EYFP-NHL25 (B) with CFP-NHL4. C-D: Validation of LRL> and FRL> as functional PTS1 of NHL6 and NHL25, respectively, where the EYFP-PTD constructs of NHL6 and NHL25 were transiently expressed in onion epidermal cells. Peroxisomes were labeled with RFP-SKL> (Matre et al., 2009). The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown, except for A (n=1). Expression times are 18 h for C-D and 1 week for A-B.

Four additional NHL homologs (Table 3.3) were predicted to carry PTS1s (NHL39, NHL39H1, NHL13H1 and NHLx). The full-length proteins were also fused N-terminally with EYFP. Notably, *NHL13H1* cDNA was not available and was cloned from genomic DNA (see 2.2.2.1). All fusion proteins were targeted to organelle-like structures upon transient expression in onion epidermal cells (Figure 3.13) and showed similar localization patterns as NHL4, NHL6 and NHL25 in the same expression system. The aggregate like-structures were more pronounced for fusion proteins of NHL13H1 and NHL39H1 (Figure 3.13, C and D). Trials to identify the EYFP-labeled structures as peroxisomes failed in onion epidermal cells, and need further investigation. To sum up, all NHLs tested were targeted to organelle-like structures in onion epidermal cells and should be tested in other expression system in order to identify the identity of these subcellular structures.

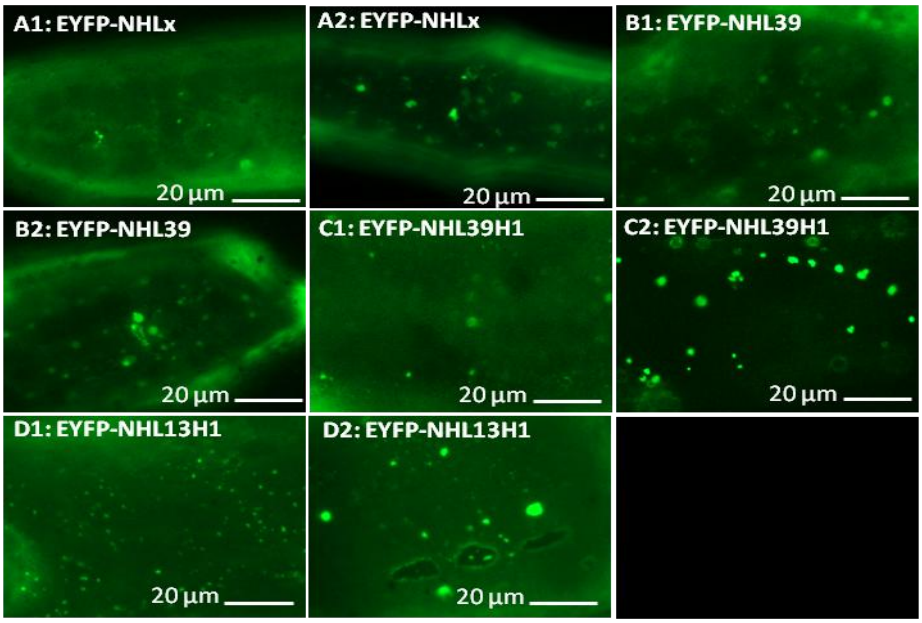


Figure 3.13: *In vivo* subcellular targeting of additional NHL proteins
The full-length proteins of NHL39, NHL39H1, NHL13H1 and NHLx were fused N-terminally with EYFP. The fusion proteins were targeted to organelle-like structures, upon transient expression in onion epidermal cells. Images X2 show formation of aggregate-like structures in different cells for NHL proteins. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown. Expression times (18 h for images X2 and 1 week for images, X1).

3.4.2 Isolation of homozygous *nhl4*, *nhl6*, and *nhl25* mutants

To initiate molecular analyses, homozygous mutants from *Arabidopsis* T-DNA insertion lines were isolated for five NHL proteins (see 2.2.1.5). Three homozygous mutants (*nhl4*, Sail_681_E12; *nhl6*, SALK_148523; *nhl25*, SALK_113216) will be mainly represented in this study (Figure 3.14, A). The T-DNAs were located in the 300-untranslated regions (UTR) in *nhl4*, in the 1st of 2 exons in *nhl6*, and in the 2nd of 2 exons in *nhl25*. In order to obtain homozygous plants, a series of genomic PCRs were applied using the two gene-specific primers (LP and RP) together with the T-DNA specific primer (LBa1: SALK or LB1S: SAIL). Several homozygous plants were identified for each line (Figure 3.14, B). Next, after seeds collection and growth of the next generation, one representative homozygous mutant for each line was verified by applying genomic PCR using either two gene-specific primers (LP and RP, for the wt allele) to confirm the absence of any wild-type allele or (LBa1 or LB1S and RP) to confirm the presence of the T-DNA insertion (Figure 3.14, C).

In preliminary phenotypic analyses, both *nhl6* and *nhl4* showed a dramatic developmental phenotype. On MS plates containing 3% sucrose, homozygous *nhl4* mutants were chlorotic, dwarfed, highly retarded in growth, and accumulated anthocyanins. All dwarf plants died (e.g., Figure 3.14, D1), but a few chlorotic plants recovered after transfer to soil and slowly developed true leaves, inflorescences and a few seeds (e.g., Figure 3.14, E1). Homozygous *nhl6* mutants were highly retarded on MS plates (Figure 3.14, F1) but grew at normal speed and were indistinguishable from wt plants after transfer to soil. For unknown reasons, the *nhl4* phenotype was less pronounced in the new generation, but still obvious for 20-30% of the homozygous plants. Moreover, *nhl6* retardation of growth was no longer observed in the next generation. The variation in phenotypes needs to be further analyzed in the future.

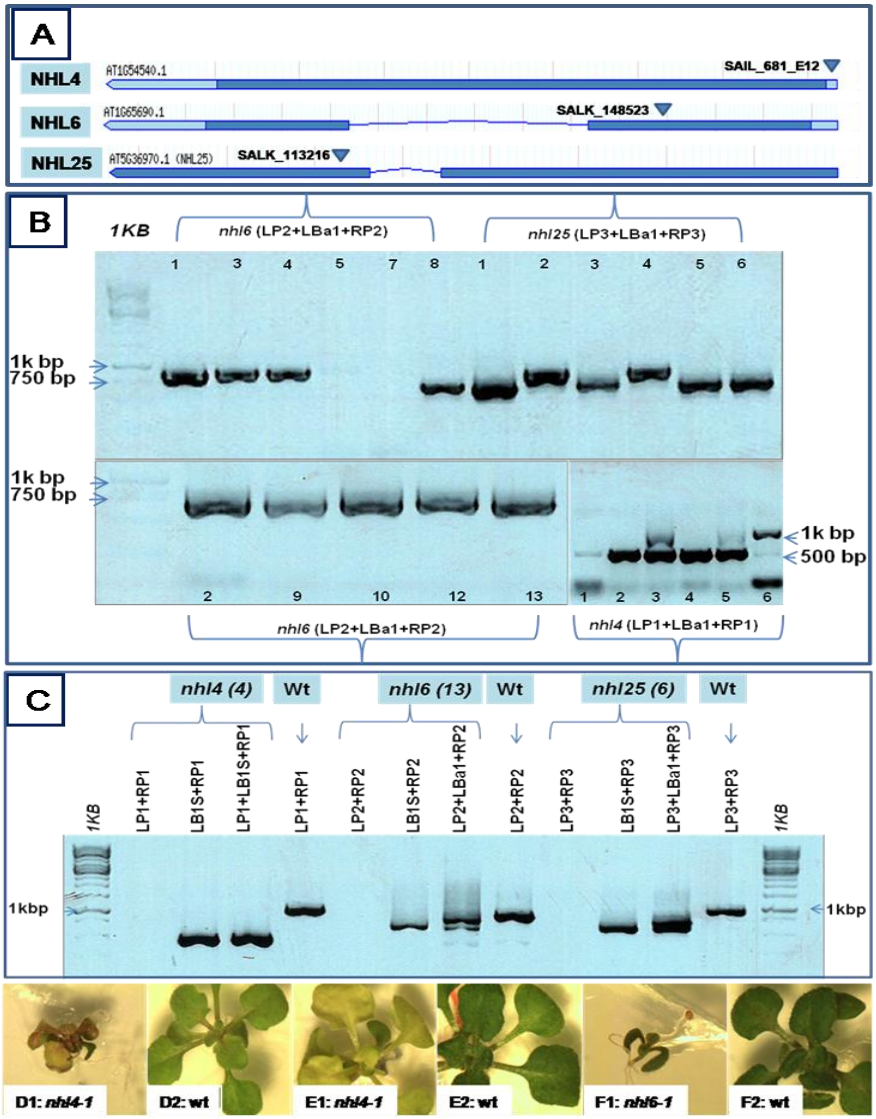


Figure 3.14: Isolation of *nhl* mutants by genomic PCR
A: Diagram of *Arabidopsis NHL4*, *NHL6* and *NHL25* genes. B: indicates the identified homozygous mutants by genomic PCR for *nhl4* (1, 2, and 4); nhl6 (2, 8, 9, 10, 12, and 13); nhl25 (1, 3, 5, and 6). The wt band for all *NHL*s is (=~ 1 kb), while the T-DNA specific band size for all *NHL*s is (=~400-700). C: Confirmation of homozygous mutant representatives for each gene by different combination of primers as shown in the figure. D-F: Developmental defect of *Arabidopsis* mutants deficient in *NHL4* and *NHL6* (single experiment). Plant images were taken at the age of 19 (D, F) and 26 days. Mutants are shown magnified.

3.4.3 Generation of *NHL* overexpresser and amiRNA lines

To study NHL4, NHL6 and NHL25 function more specifically, stable *Arabidopsis* lines with specific gene overexpression or knockdown by amiRNAs were generated. Both *NHL4* and *NHL25* stable overexpresser lines and amiRNA lines were generated, while delayed for *NHL6* because of cloning and subcloning difficulties (Table 3.4). The available lines were produced by standard procedures, i.e. subcloning the target genes into binary vectors (see 2.2.2.1) and further plant transformation (see 2.2.1.4). After obtaining the transformed seeds, T1 plants were selectively (see 2.2.1.4) isolated, and their seeds were subsequently harvested. The available T1s for each line were approved as a preliminary step by genotyping of the plant genomic DNA for the presence of the transformed constructs using gene- and vector-specific primers (see 2.2.1.4). Moreover, the full-length *NHL4* was fused in the back of *EYFP*, and the fusion construct is available in pGEMT Easy vector. The *EYFP-NHL4* will be subcloned into binary vector for stable and transient expressions in order to study the subcellular localization of NHL4 in plant tissues.

Table 3.4: List of *NHL* overexpresser and amiRNA lines
Transformed seeds availability is indicated by (+), and the Transformation (TF) rate is indicated. T1 available lines number is indicated from the successfully genotyped plants. TF rates, expressed as 'percentage transformation', were calculated as [(#marker-resistant seedlings)/(total # seedlings tested)] x 100 (Clough and Bent, 1998).

No	Gene/ amiRNA	Plasmid/ promoter	TF Seeds	TF Rate (%)	T1 lines
1	*NHL4*	pBA002/35S	+	30	11
2	*NHL4*	pER10/Estradiol	+	0.28	7
3	*EYFP-NHL4*	subcloned in pGEMT			
4	*NHL25*	pBA002/35S	+	30	4
5	*NHL25*	pER10/Estradiol	+	30	1
6	*NHL6*	cloning delayed			
7	*EYFP-NHL6*	cloning delayed			
8	amiRNA (*NHL4*)	pER10/Estradiol	+	30	5
9	amiRNA (*NHL25*)	pER10/Estradiol	+	0.25	5
10	amiRNA (*NHL6*)	cloning failed	+		
11	control	pBA002/35S	+		
12	control	pER10/Estradiol	+	0.2	3

3.4.4 Plant immunity assays

3.4.4.1 *Pst* DC3000 proliferation in *Arabidopsis*

One crucial characteristic of resistant plants is their ability to restrict *in planta* growth of avirulent bacteria. Virulent pathogens (e.g., *Pst* DC3000) inoculated at low concentrations (e.g., 10^4 CFU/cm^2 leaf tissue, which approximately corresponds to an inoculation of 10^6 CFU/ml) can colonize the host tissue and multiply more than 10,000-fold within the host tissue in several days (up to 10^8 CFU/cm^2 leaf tissue) (Katagiri et al., 2002). In contrast, nonpathogenic mutant strains (e.g., *Pst* DC3000 hrpH$^-$ mutant, deficient in TTSS secretion system) or avirulent pathogens (e.g., *Pst* DC3000 carrying avrRpm1 or avrRpt2 effectors) in the same time course will either not multiply significantly or grow only 10- to 100-fold within the host tissue [see 2.1.2.3 and 1.2.1.2, (Katagiri et al., 2002)]. The assay was established in the group by monitoring the growth of virulent *Pst* DC3000 and avirulent *Pst* DC3000 (avrRpt2) on wt Col-0 after syringae infiltration of 10^6 CFU/ml. The virulent bacteria proliferated in the wt up to 10,000 fold in 2 days, while the avirulent strain only proliferated 10 fold (Figure 3.15, A). Moreover, avirulent bacteria produced no disease symptoms, while virulent bacteria caused chlorosis and necrosis of the infiltrated tissue of a susceptible host plant within 3-4 days (data not shown). From these data, and consistent with previous literature, wt Col-0 was more resistant to the avirulent than the virulent strain. The above mentioned observations were considered to be successful and nicely reproduced two times similarly and aligned with the published data (Katagiri et al., 2002).

3.4.4.2 Proliferation of avirulent *Pst* DC3000 in *nhl* mutants

To investigate innate immunity in *nhl4, nhl6, and nhl25* mutants, *Pst* DC3000 (avrRpt2) growth was monitored in leaves of intact plants (see 2.2.4.1) and compared with wt Col-0, *ndr1-1*, see 1.2.3.1, (Century et al., 1995; Century et al., 1997) and *npr1-1*, see 1.2.2.3 (Cao et al., 1994) (Figure 3.15, C). wt Col-0 is resistant to the avirulent strain because of ETI (see 1.2.1.2), while *ndr1.1* and *npr1.1* are susceptible plants because of the loss of NDR1 and NPR1, respectively (see 1.2.2.3 and 1.2.3.1). The population of *Pst* DC3000 (avrRpt2) in wt Col-0 plants (2-4 days after inoculation) proliferated only 10 fold, while in *ndr1-1* and *npr1-1* mutants the bacteria proliferated 10,000 and 1000 fold, respectively. Interestingly, the population proliferated around

1000 fold in all of the 3 *nhl* mutants (Figure 3.15, C). Disease symptoms appeared on the inoculated leaves of all mutants, 2-4 days after inoculation while wt Col-0 did not show any symptoms (Figure 3.15, C3). In summary, consistent with previous literature both *ndr1.1* and *npr1.1* plants were more susceptible than wt Col-0, besides that, *nhl* mutants were more susceptible, which indicates their probability to be important in pathogen resistance. However, it needs to be pointed out that the proliferation assay remained preliminary due to insufficient number of only two biological replicates in independent experiments. Several trials were done to generate the 3rd repetition but these were hindered by some technical difficulties with the available plant growth chamber facility.

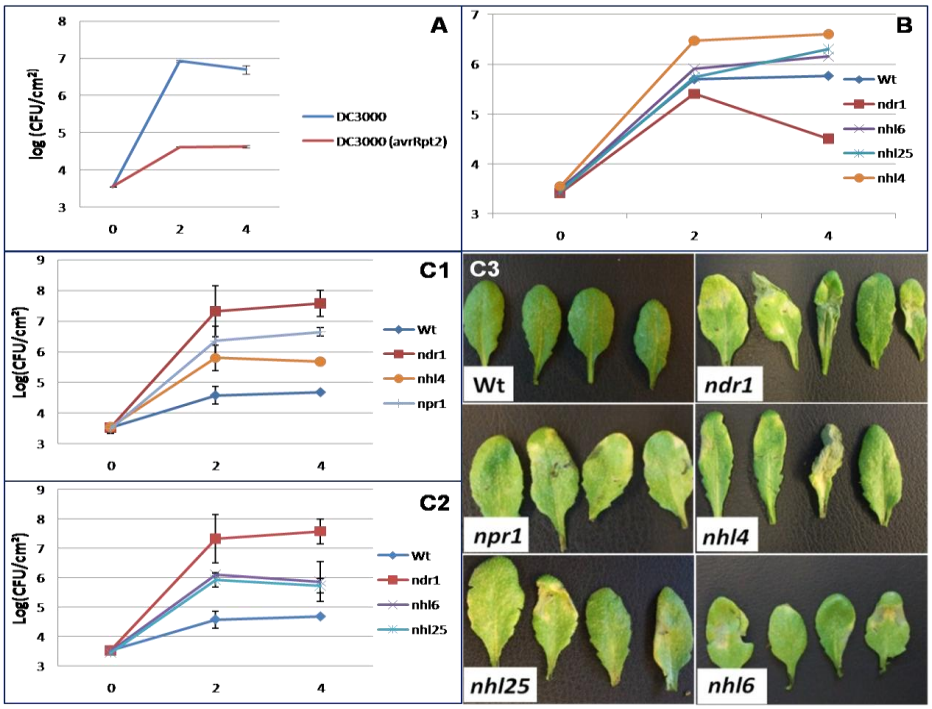


Figure 3.15: Pathogen proliferation analyses in *nhl* mutants
A: *Pst* DC3000 proliferation comparison in wt Col-0 between virulent and avirulent *Pst* DC3000 (avrRpt2) (n=2 with similar results, and SD between the 2 experiments is shown). B: Pathogen proliferation analyses (virulent *Pst* DC3000) in *nhl* mutants (n=1). C: Pathogen proliferation analyses by [*Pst* DC3000 (avrRpt2)] (left) and increased pathogen susceptibility (right panel) in *nhl* mutants (n=2). For A-C, mature soil-grown plants were infiltrated with low density avirulent *Pst* DC3000 (avrRpt2) or virulent *Pst* DC3000 solutions (concentration of 10^6 CFU/ml). The horizontal axis is in days. Bars show the SD, for calculations see 2.2.4.1. C: These results were obtained twice with a higher SD for some readings, while the 3rd repetition failed twice because of plant growth technical problems and needs to be further investigated.

3.4.4.3 Proliferation of virulent *Pst* DC3000 in *nhl* mutants

Pst DC3000 bacteria were used in a single experiment to address their growth in *nhl4, nhl6, and nhl25* mutant plants and to compare this with avirulent bacterial growth (see 3.4.4.2). *Pst* DC3000 growth was monitored in leaves of intact plants and compared with wt Col-0, *ndr1-1* ((Century et al., 1995; Century et al., 1997), Figure 3.15, B). The population of *Pst* DC3000 in wt Col-0 plants and *ndr1-1* mutants (2-4 days after inoculation) proliferated 1000 fold. Moreover in *nhl* mutants the bacteria proliferated similar to wt Col-0 (Figure 3.15, B).

3.4.4.4 Callose deposition analysis in *nhl* mutants

Furthermore, callose deposition induced by flg22 (see 1.2.1.1 and 2.2.4.2) was further investigated in *nhl* mutants. Two *Arabidopsis* wt ecotypes (Col-0 and Ws-0), *pen2-1* (Lipka et al., 2005; Clay et al., 2009), *nhl4*, *nhl6*, and *nhl25* were treated by 1 µM flg22 (see 2.2.4.2). As expected, callose was deposited in wt Col-0, while Ws-0 showed approximately no callose depositions consistent with being a negative control (Gomez-Gomez et al., 1999). Moreover, *pen2-1* and *nhl* mutants show callose depositions which quantitvely varied between plants in each of the mutants (Figure 3.16). The number of callose deposits per microscopic field was calculated from leaves from independent plants, by using the ImageJ software (Figure 3.16), and the average number of callose deposits was calculated and blotted for each mutant (Figure 3.16). The callose deposits average number in *nhl* mutants and *pen2-1* demonstrates that the callose depositions decreased as compared to wt Col-0, as averaged from two biological dublicates.

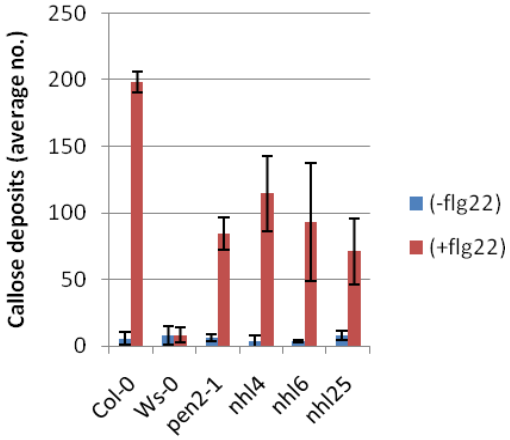


Figure 3.16: Callose depositions analysis in *nhl* mutants
Wt (Col-0 and WS-0), *pen2-1*, *nhl4*, *nhl6* and *nhl25* seedlings were incubated in the presence or abscence of 1 µM flg22 (see 2.2.4.2), seedlings were then stained by aniline blue and callose deposits were detected by fluorescence microscopy (see 2.1.4.2). The graph shows average number of callose deposits of 5 different leaf samples from at least 5 independent seedlings. Callose deposits were analyzed using ImageJ. This experiement was repeated twice in 3 replicates (n=2x3), and the bars indicate SD, for calculations see 2.2.4.2.

3.4.5 Analysis of metabolic peroxisome functions in *nhl* mutants

To investigate whether NHL proteins of interest indirectly participate in fatty acid β-oxidation (see 1.1.1.1) and IBA-to-auxin conversion, the homozygous T-DNA mutants (*nhl4*, *nhl6* and *nhl25*) were subjected to sucrose dependence (see 2.2.5.1) and IBA-response (see 2.2.5.2) assays using the *pex14* null mutant as a positive control (Orth et al., 2007). In the absence of sucrose, hypocotyl elongations of the *nhl* mutant seedlings was slightly inhibited similar to wt plants, while hypocotyl length was significantly reduced in the *pex14* mutant, consistent with its defect in fatty acid β-oxidation (Figure 3.17, A). This growth inhibition was largely rescued by exogenous sucrose. Second, the response of the *nhl* mutants to IBA was analyzed. Low levels of IBA (10-15 µM) inhibited root elongation in wt and *nhl* mutants similarly, while *pex14* mutant was largely insensitive to IBA, consistent with previous reports [(Zhang and Hu, 2010), Figure 3.17, B]. These observations indicate that it is likely that NHL proteins are not involved in fatty acid β-oxidation.

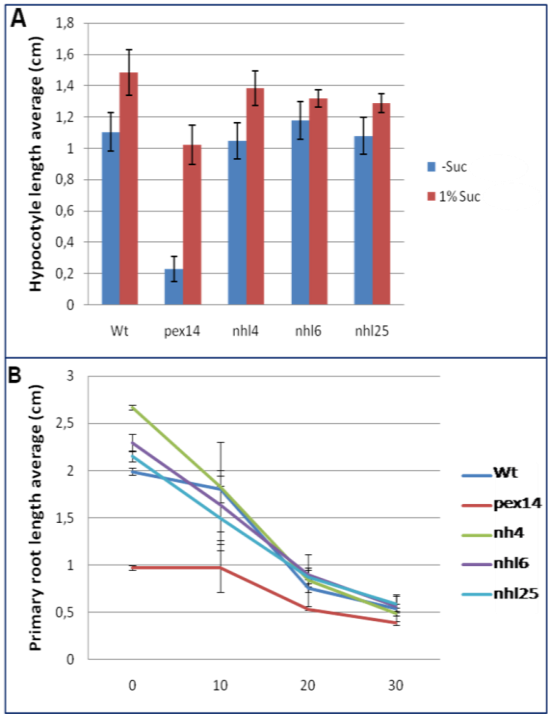


Figure 3.17: Analysis of metabolic peroxisome functions in *nhl* mutants
A: Sucrose dependence assay. Hypocotyl lengths of seedlings grown for 6 d in the dark on half-strength LS media with or without the supplement of 1% sucrose (w/v) are shown. The experiment was repeated 3 times with similar results (n=3). B: Effect of IBA on primary root elongation. Plants were grown for 7 d in the light on half-strength LS media supplemented with 0, 10, 20, 30 µM IBA. The experiment was repeated 2 times with similar results (n=2). Hypocotyl and root lengths were measured by Image J. Bars indicating SD. For SD calculations see 2.2.5.

3.4.6 Expression analysis of *NHL* genes

According to Genevestigator, expression of *NHL4* and *NHL6* was analyzed. *NHL4* transcripts were constitutively expressed in seedlings and roots, while *NHL6* transcripts were mainly found in senescent leaves (Figure 3.18, A). *NHL6* appeared to be induced by several infections (virulent and avirulent bacteria, fungi, oomycetes and viruses). *NHL6* was also induced when treated by different types of elicitors e.g. flg22. In contrast, *NHL4* was induced by bacteria, GST-NPP1, SA and ABA but *NHL4* was less pronounced than *NHL6* (Figure 3.18, C). *NHL25* was induced by avirulent *Pst* DC3000 infection of wt Col-0, that harbors one of the effectors (avrRpm1, avrRpt2, avrB, or avrRps4). Additionally, *NHL25* was induced by SA, while it was not induced by either ethylene or JA (see 1.2.3.1, (Varet et al., 2002)). In summary, the three *Arabidopsis NHL* genes appeared to be induced in response to several biotic stresses. These expression patterns are supporting their suggested importance in plant resistance to pathogen infection.

Real time PCR (see 2.2.2.8) was used for quantification of mRNA transcripts levels to investigate *NHL* genes induction. wt Col-0 plants were grown and treated either by flg22 or pathogen (see 2.2.4). RNA was then isolated from the treated leaves, and subjected to real time PCR (for primer optimization and testing of the genes, see 2.2.2.8). Preliminary expression analyses indicated pathogen-dependent mRNA accumulation for *NHL*s (Figure 3.19). *NHL4*, *NHL6* and *NHL25* transcripts accumulated similar to *PR2* (Edreva, 2005) after 8 h post infection (P.I.) by virulent *Pst* DC3000 (Figure 3.19, A). *NHL6* transcripts specifically and to lesser extent *NHL25* accumulated 8 h P.I. when the bacteria carried the effector avrRpt2 (Figure 3.19, B). Specifically, *NHL6* transcripts accumulated after treatment with flg22 (Figure 3.19, C), but did not accumulate in *fls2* plants carrying mutations in the flagellin receptor gene *FLS2* (Zipfel et al., 2004; Heese et al., 2007) suggesting its role in PTI (see 1.2.1.1 and Figure 3.19, D). Furthermore, *NHL6* induction was not affected in *npr1.1* plants (see 1.2.2.3) after treatment by flg22 which shows that *NHL6* induction is NPR1-independent, i.e., not induced downstream of NPR1 (Figure 3.19, D). However, these data are preliminary, but indicated the importance of NHL6 in both PTI and ETI. Furthermore, *NHL4*, *NHL6* and *NHL25* appear to be induced by bacterial pathogens.

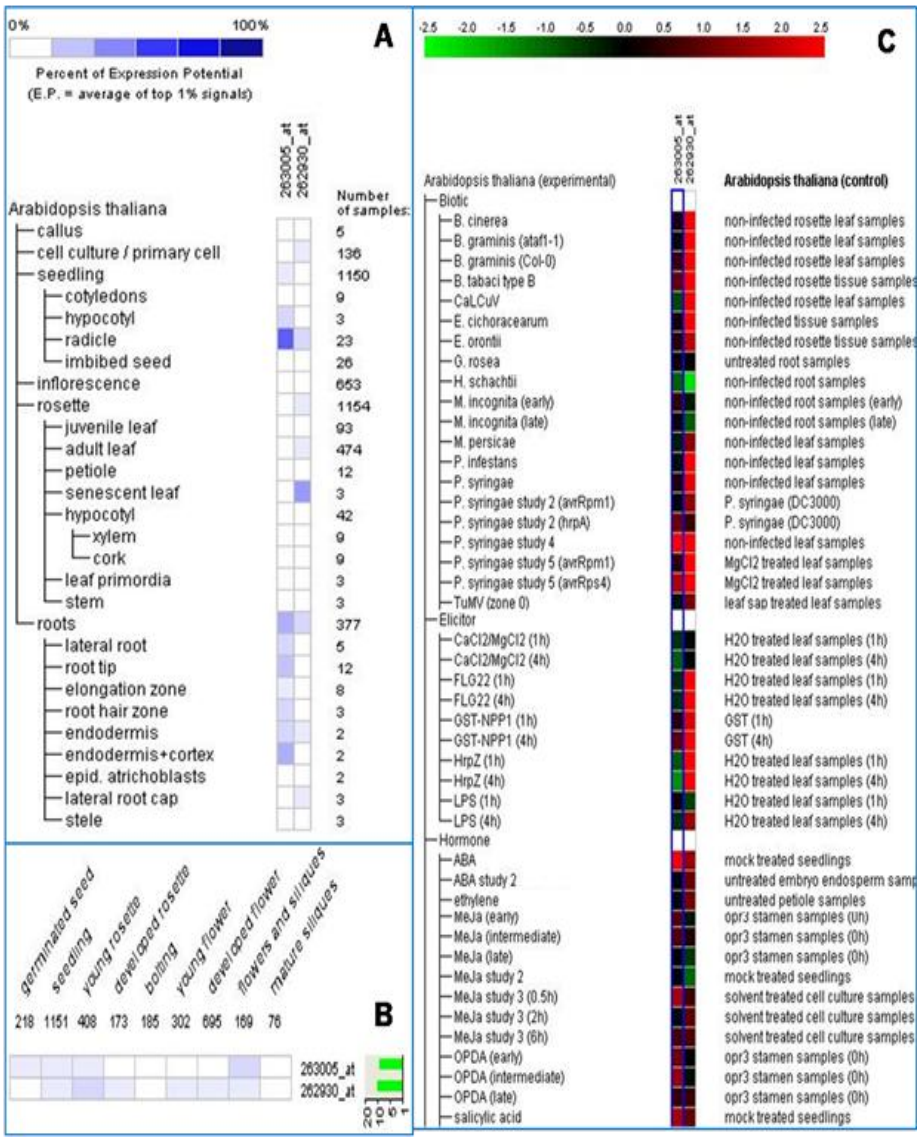


Figure 3.18: Gene expression analyses for *NHL4* and *NHL6*
Gene expression analyses of *Arabidopsis NHL4* and *NHL6*, which were investigated in the present study. A, B and C are images representing anatomy, development, and stress-related expressions, respectively. The expression data derive from microarray experiments and were retrieved using Genevestigator (www.genevestigator.com; (Zimmermann et al., 2004)). High and low expression levels are reflected semi-quantitatively by *dark* and *light* coloring, respectively. 263005_at: *NHL4*; 262930_at: *NHL6*.

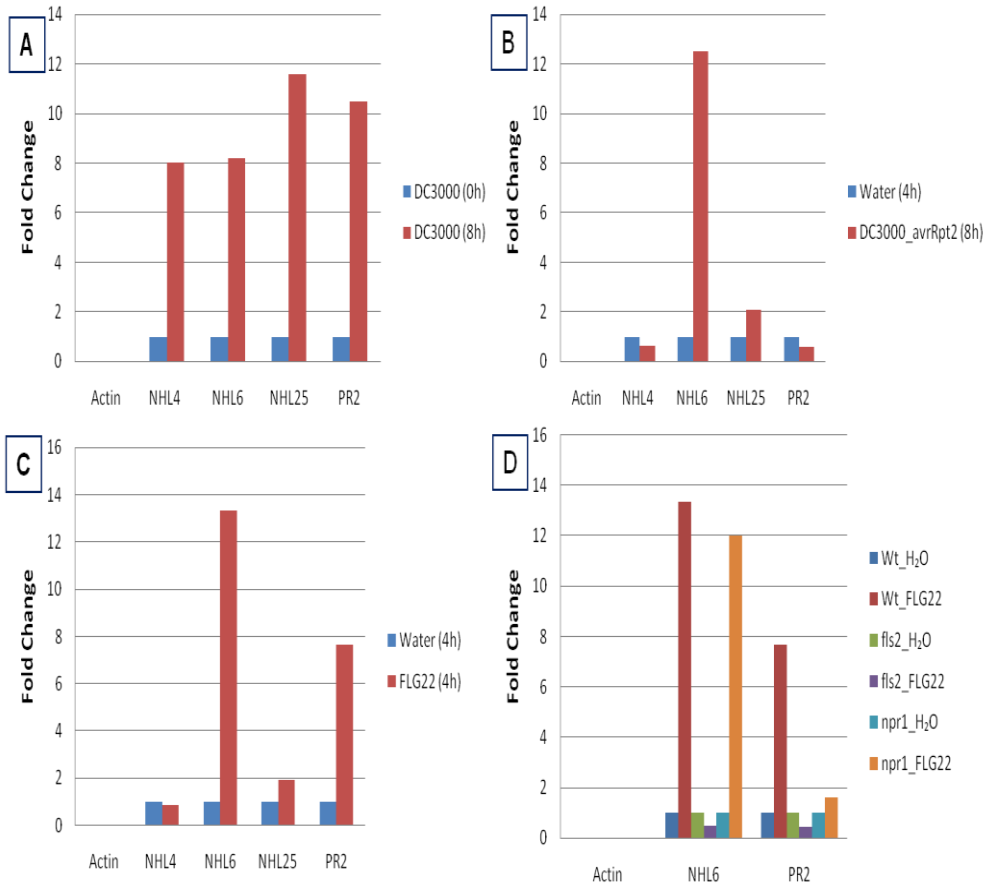


Figure 3.19: Pathogen induction of *NHL* genes
A: *NHL4*, *NHL6* and *NHL25* genes were induced by the virulent *Pst* DC3000. B and C: *NHL6* and *NHL25* genes were induced by the avirulent *Pst* DC3000 and by flg22. D: *NHL6* is expressed by flg22 in Wt Col-0 and *npr1-1* mutant but was not induced in the *fls2* mutant upon the same treatment. A: Plants infiltrated with virulent *Pst* DC3000. B: Plants infiltrated with water or *Pst* DC3000 carrying the *avrRpt2* avirulence gene. C: Plants infiltrated with water or 1 µM flg22. D: Wt Col-0, *fls2*, *npr1-1* plants infiltrated with water or 1 µM flg22. Leaf tissues were collected at the indicated time points and analyzed by real time PCR. The data are preliminary (n=1). For A-B, six-week-old soil-grown plants were infiltrated with high density virulent *Pst* DC3000 or avirulent *Pst* DC3000 (avrRpt2) solutions (concentration of 10^8 CFU/ml).

3. 5. AtIAN protein family investigations

AIG (here referred to as AtIAN) proteins (see 1.2.3.2), are a family of GTPases, one member of which (AIG1/AtIAN8) are suggested to be involved in the RPS2-dependent plant resistance pathway [(Liu et al., 2008), see ETI, 1.2.1.2] based on the expression of *AIG1/AtIAN8* in *Arabidopsis* after infection by avirulent *Pst* (avrRpt2) (Reuber and Ausubel, 1996). Apart from two studies in *Arabidopsis* (Reuber and Ausubel, 1996; Liu et al., 2008), AtIANs were not reported to be further studied. AtIAN12 (At4g09940) had been identified by experimental proteomics in *Arabidopsis* leaf peroxisomes (Reumann, unpub. data). AtIAN12 terminates with IIM>, which resembles plant PTS1 tripeptides such as AKM>. However, PTS1 prediction algorithms did predict neither AtIAN12 nor any of its homologs as peroxisome-targeted PTS1 proteins. However, several family members were scored slightly below threshold in the gray zone in which several true positive peroxisomal PTS proteins are found ((Lingner et al., 2011), e.g., AtIAN3, and AtIAN8, Table 3.5).

Table 3.5: PTS1 protein prediction scores for AtIAN homologs
The threshold of the prediction scores for predicted peroxisome targeting for the PWM model is 0.412 (Lingner et al., 2011). The gray zone is up to 0.130.

Acronym	AGI code	C-term. tri-peptide	PWM model score
AtIAN3	At1g33890	SIL>	0.326
AIG1/AtIAN8	At1g33960	SIL>	0.216
AtIAN1	At1g33830	VKL>	0.128
AtIAN11	At4g09930	IIL>	-0.47
AtIAN12	At4g09940	IIM>	-0.52

3.5.1 *In vivo* subcellular localization of AtIAN proteins

To validate peroxisome targeting of AtIAN12, the full-length protein was fused N-terminally with EYFP under the control of CaMV 35S promoter in two different vectors (pCAT and pBA002; see 2.1.3). The pBA002 binary vector was used for *Agrobacterium*-mediated transformation of tobacco intact leaves (see 2.2.3.2). The vector pCAT was used in both onion epidermal cells (see 2.2.3.1) and tobacco protoplast (see 2.2.3.3) transformations. Upon transient expression in onion epidermal cells the fusion protein was detected in morphologically diverse subcellular structures, including

interconnected punctate structures ("beads on a string", Figure 3.20, A-C, 18-48 h expression times). Some EYFP-labeled subcellular structures were demonstrated to coincide with CFP-labeled peroxisomes in a limited number of cells in two different experiments (Figure 3.20, A and C, 18-48 h expression times). In further experiments, however, the EYFP-labeled subcellular structures, even though intensively ($n \geq 6$) investigated, no longer coincided with CFP-labeled peroxisomes for unknown reasons (Figure 3.20, B and C). EYFP-AtIAN12 was also transiently expressed in tobacco protoplasts. As in onions, yellow fluorescence was detected in organelle-like structures and sometimes interconnected punctate structures surrounding plastids (Figure 3.21, A-C). However, both subcellular structures did not coincide with CFP-labeled peroxisomes (data not shown). EYFP-AtIAN12 was also co-expressed with an ER marker [(OFP-ER, see 2.1.3, (Frank et al., 2008)]. Preliminary data showed partial co-localization of EYFP-AtIAN12 with the ER marker (Figure 3.21, D and E). The data imply that AtIAN12 partially or transiently localizes to the ER.

To further analyze the subcellular localization of AtIAN12, EYFP-AtIAN12 was transiently co-expressed with CFP-PTS1 [see 2.1.3 (Zhang and Hu, 2008)] in tobacco leaves by *Agrobacterium*-mediated transformation (see 2.2.3.2). EYFP-AtIAN12 was detected in organelle-like structures that mostly coincided with CFP-labeled peroxisomes (Figure 3.22, A-C). However in some leaf cells, the coincidence was only partial or even absent (data not shown). EYFP-AtIAN12 was also co-expressed with a CFP-fused markers for the ER and Golgi [see 2.1.3 (Nelson et al., 2007)]. Additionally, mitochondria were stained by incubation of the leave tissue for 1 h in the red stain (1 µM MitroTracker red-CMXRos, Invitrogen, USA). In the three cases, no co-localization was detected, indicating that AtIAN12 is not targeted to ER, Golgi, or mitochondria in this expression system (Figure 3.22, C-E). Non-punctate interconnected structures as observed in onion epidermal cells and tobacco protoplasts were not seen in tobacco leaves. In summary, it was concluded from the subcellular targeting data for full-length AtIAN12 in onions, tobacco protoplasts and intact tobacco leaves that AtIAN12 is targeted to peroxisomes in intact tobacco leaves, possibly due to special "defense conditions" caused by *Agrobacterium*-mediated transformation.

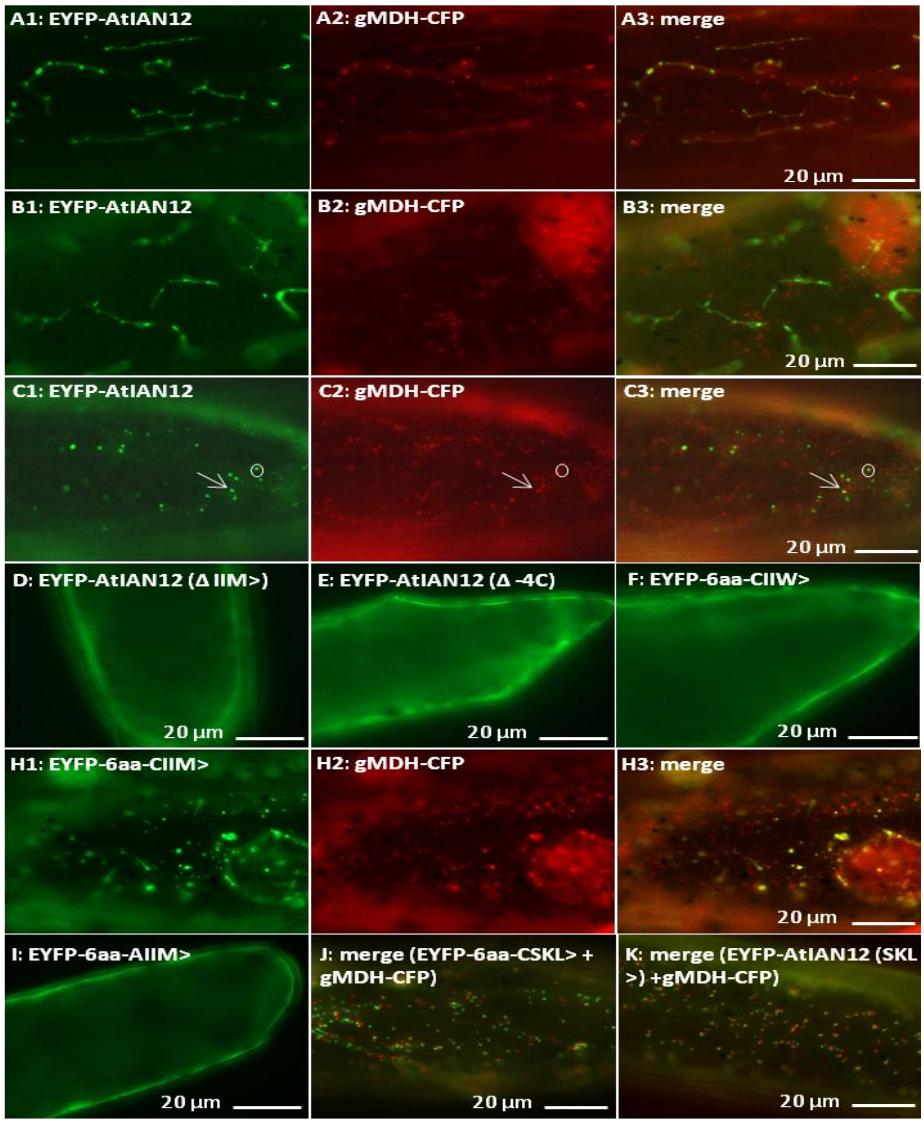


Figure 3.20: *In vivo* subcellular localization of AtIAN12 in onions
A-C: EYFP-AtAIN12 was targeted to organelle-like structures and interconnected structures that partially coincided with CFP labeled peroxisomes upon transient expression in onion epidermal cells. H: EYFP-PTD (AtIAN12, IIM>) shows enlarged vesicle structures. D-F and I: SDM of the predicted isoprenylation motif (CIIM>) in both EYFP-AtIAN12 and EYFP-PTD (Table 3.6) made the proteins to remain cytosolic. J-K: replacing IIM> with SKL> targeted the full-length protein and decapeptide fusions to peroxisomes. Peroxisomes were labeled with gMDH-CFP. The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 are shown. Expression times (18 h for A, D-K and1 week for B-C).

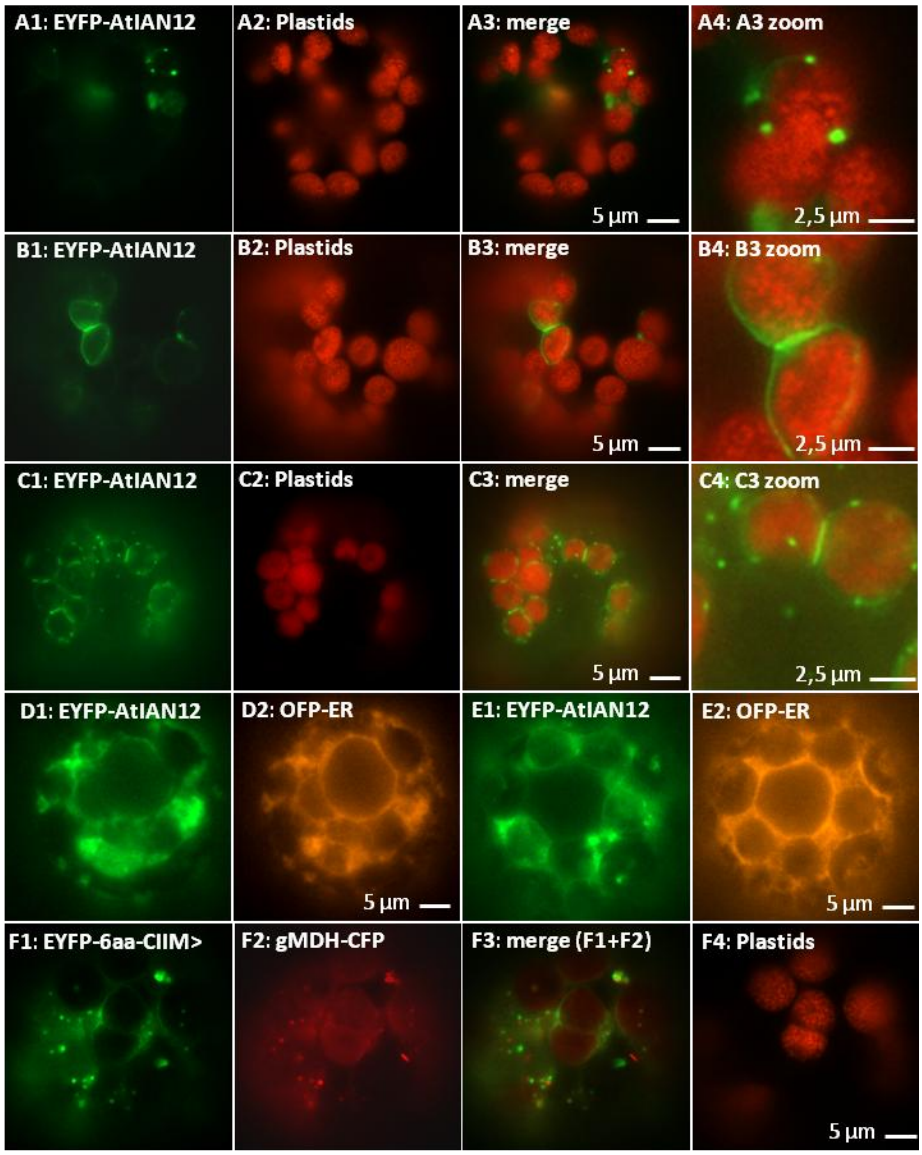


Figure 3.21: *In vivo* subcellular localization of AtIAN12 in tobacco protoplasts
A-C: EYFP-AtIAN12 was targeted to organelle-like structures, and interconnected structures surrounding plastids. D and E: preliminary data show partial localization of AtIAN12 to the ER. F: EYFP-decapeptide (AtIAN12, IIM>) shows vesicle (atypical) structures. Peroxisomes were labeled with gMDH-CFP (Fulda et al., 2002). The ER was labeled by OFP-ER (Frank et al., 2008). The cyan fluorescence was converted to red. For fluorescence image acquisition details, see 2.1.4.1. Representative images of reproducible results obtained ≥3 except for D-F (n=1) are shown. Expression times (18 h-48 h).

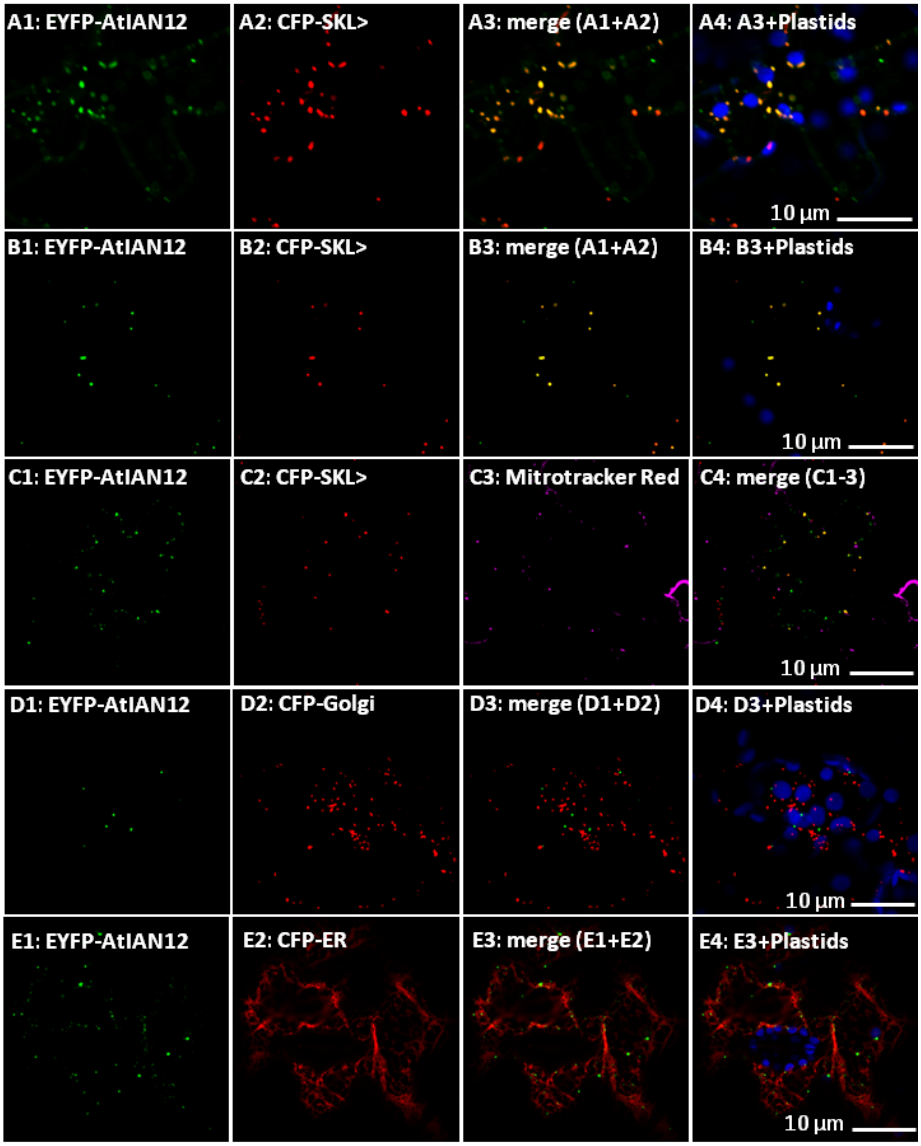


Figure 3.22: *In vivo* subcellular localization of AtIAN12 in tobacco leaves
Confocal laser scanning microscopic images (CLSM) for leaf cells from 6-week-old intact tobacco leaves. A-C: EYFP-IAN12 was targeted to peroxisomes labeled with CFP–PTS1. D and E: EYFP-AtIAN12 did not coincide with CFP-Golgi or CFP-ER. C: AtIAN12 did not coincide with mitochondria that were stained by incubating the leaf tissue for 1 hour in 1 µM Mitrotracker red solution (Invitrogen, USA). In A-E red signals indicate CFP; blue signals indicate plastids; pink signals indicate MitoTracker-stained mitochondria. For organelle marker details see 2.1.3. For fluorescence image acquisition details, see 2.1.4.4. Representative images of reproducible results obtained ≥3 are shown. Expression times (2-7 d).

In order to study if IIM> is a functional PTS1, the EYFP-PTD of AtIAN12 was constructed. Upon transient expression in onion epidermal cells, the fusion protein targeted to both organelle-like and larger vesicle-like structures that reproducibly and convincingly did not coincide with CFP-labeled peroxisomes (Figure 3.20, H). Similar results were obtained in a single experiment done with tobacco protoplasts (Figure 3.21, F). Thus, the data indicated the EYFP-PTD of AtIAN12 targeting to non-peroxisomal subcellular structures.

Peroxisome targeting for AtIAN12 (Figure 3.22, A-D) and the possibility that AtIAN12 carried an atypical PTS1 tripeptide prompted us to investigate subcellular targeting of additional AtIAN family members with PTS1 prediction scores higher than that of AtIAN12, including the prototypical family member, AtIAN8/AIG1. Two full-length proteins (AtIAN8, SIL> and AtIAN11, IIL>, see Table 3.5) were fused N-terminally with EYFP. The reporter fusion proteins did reproducibly targeted to organelle-like structures in onion epidermal cells (Figure 3.23, A-C, n≥3, 18 h to 1 week) and tobacco protoplasts (Figure 3.23, D-G, n=2 , 18-48 h). Simultaneously with the full-length fusions AtIAN8 and AtIAN11, their proposed PTDs were fused N-terminally with EYFP. The reporter fused PTDs of both proteins targeted weakly to organelle-like structures in both onion epidermal cells (Figure 3.23, L and M, n≥3, 18 h- 1 week) and tobacco protoplasts (Figure 3.23, J and K, n=1, 18-48 h). The identified structures did not coincide with CFP-labeled peroxisomes in both systems. However, despite variation of expression times and a significant number of experimental repetitions, these organelle-like structures could not be identified as peroxisomes. Identification of the nature of these non-peroxisomal organelle-like structures was beyond the scope of this study and requires further investigations.

In conclusion, by using the transient expression systems of onion epidermal cells and tobacco protoplasts, AtIAN8, AtIAN11 and AtIAN12 and their reporter fused PTDs were targeted mainly to unidentified non-peroxisomal organelle-like structures. However, these data could not validate peroxisome targeting, but indicated that the targeting signal is located in the C-terminal decapeptides.

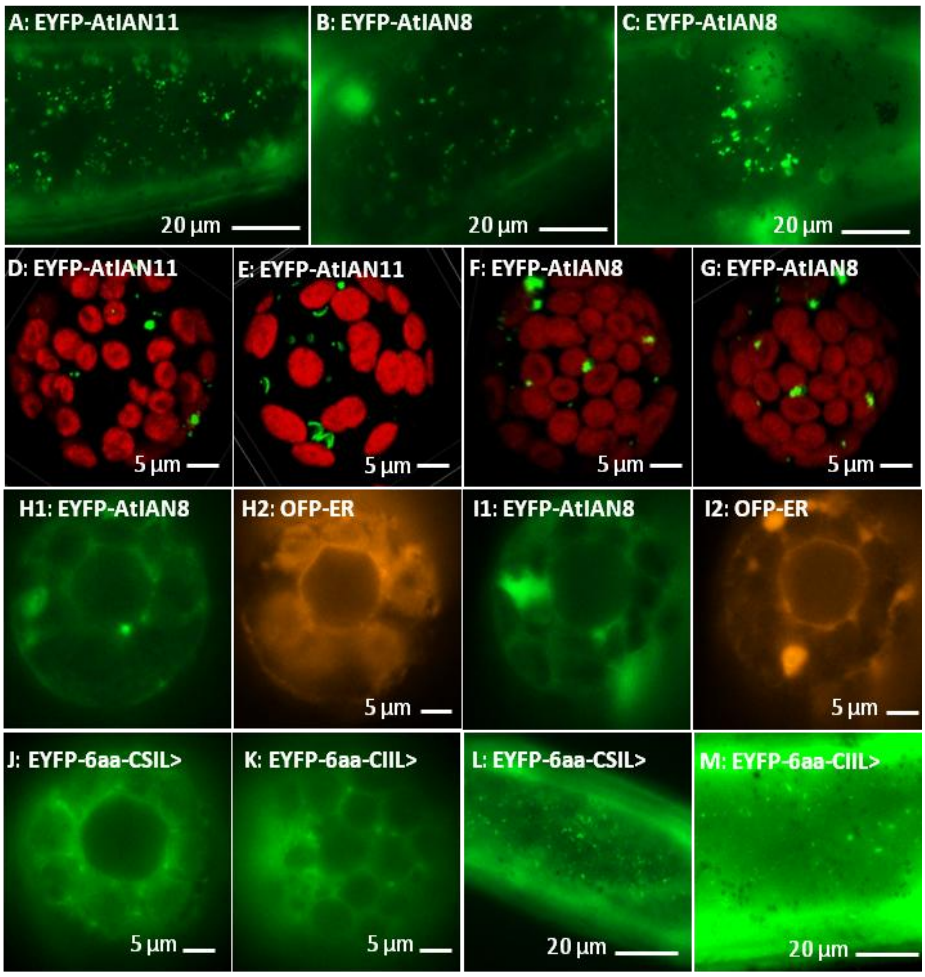


Figure 3.23: *In vivo* subcellular localization of AtIAN8 and AtIAN11
The EYFP-AtIAN8 and EYFP-AtIAN11 were transiently expressed in isolated tobacco protoplasts and onion epidermal cells. A-C: EYFP-AtIAN8 and EYFP-IAN11 were targeted to yet unidentified organelle-like structures in onions. D-G: 3D CLSM snapshot images for EYFP-AtIAN11 and AtIAN8 in protoplasts, show targeting of both fusion proteins to yet unidentified organelle-like structures. H and I: Epifluorescent images show preliminary data for partial detection of EYFP-AtIAN8 in the ER. J and L: EYFP- PTD (AtIAN8) targeted to non-peroxisomal unidentified organelle-like structures in both protoplasts and onions. K and M: EYFP-PTD (AtIAN11) targeted to non-peroxisomal unidentified organelle-like structures in both protoplasts and onions. For organelle marker details see 2.1.3. For fluorescence image acquisition details, see 2.1.4.1, and 2.1.4.3 for D-G images. Representative images of reproducible results obtained ≥3 except for D-G (n=2) and H-K (n=1) are shown. Expression times (18-48 h for protoplast and 18 h-1 week (for L-M) for onion epidermal cells).

3.5.2 AtIAN12 appears to be post-translationally modified

To investigate whether the C-terminal decapeptides of AtIAN12, AtIAN11, and AtIAN8 might contain alternative targeting signals for subcellular organelles other than PTS1s, the full-length proteins were subjected to prediction analysis of post-translational modifications and subcellular targeting (Prenylation Prediction Suite, PrePS, http://mendel.imp.ac.at/sat/PrePS/). The three AtIAN homologs of interest were found to carry isoprenylation/farnesylation sites predicted with high probability [(Maurer-Stroh and Eisenhaber, 2005), e.g., for AtIAN12; Figure 3.24)]. Protein isoprenylation refers to the covalent attachment of a 15-carbon farnesyl or 20-carbon geranylgeranyl moiety to a cysteine residue at/or near the carboxyl terminus (Crowell and Huizinga, 2009). The isoprenylation motif is CaaX, and is located at the extreme C-terminus, where "C" is cysteine, "a" is an aliphatic residue, and "X" is usually methionine, glutamine, serine, alanine, or cysteine in case of farnesylation, and leucine or isoleucine in case of isoprenylation (Crowell and Huizinga, 2009). Protein post-translational modification starts in the cytosol (farnesyl or geranylgeranyl moiety attachment), processed in the ER (aaX cleavage followed by cysteine methylation), and further exported to its final destination from ER (Crowell, 2000; Galichet and Gruissem, 2003).

The presence of predicted isoprenylation motifs in AtIAN12 (CIIM>), AtIAN11 (CIIL>), and AtIAN11 (CSIL>) suggested that the EYFP-PTDs of the three AtIAN proteins might be targeted to and anchored in the membrane of endomembrane vesicles via attachment of an isoprenyl moiety. Thereby, the isoprenylation motif predictions overlapped with the location of possible PTS1 tripeptides (IIM>, IIL>, and SIL>). The three AtIAN proteins might be similar to PEX19 which is reported to be farnesylated [see 1.1.2, (Rucktaschel et al., 2009)]. The isoprenylation motif predicted in the C-terminus for the three IAN proteins of interest suggested that the EYFP-AtIAN (full-length and PTDs) proteins were targeted to and remained in ER-derived vesicles for isoprenylation in onion epidermal cells and tobacco protoplasts. EYFP-AtIAN12 was transported (probably via the same ER vesicles) to its final destination, mature peroxisomes, in mesophyll cells of intact tobacco leaves.

Peroxisome targeting via isoprenylation was studied in greater details representatively for AtIAN12. The predicted motif (CIIM>) was found to be conserved in its plant homologs (see Figure 3.24). Several full-length and C-terminal domain constructs containing point mutations in critical amino acid residues were constructed (Table 3.6). EYFP-AtIAN12 lacking the C-terminal tripeptide (IIM>) was no longer targeting to subcellular structures and remained cytosolic (Figure 3.20, D) in onions, indicating that the deletion of the putative PTS1 tripeptide and/or disruption of the predicted isoprenylation motif (CIIM>) prevented protein targeting to subcellular structures. Likewise, EYFP-AtIAN12 (GIIM>) (i.e., C-to-G point mutation in the predicted isoprenylation motif) remained cytosolic (Figure 3.20, E). SDMs introduced into the EYFP-decapeptide (EYFP-6aa-CIIM>) further supported the idea that C-terminal protein isoprenylation determined subcellular targeting. For example, when M at pos. -1 was mutated to W (EYFP-6aa-CIIW>), the fusion protein remained in the cytosol (Figure 3.20, F), consistent with a significantly lowered prediction score for isoprenylation (from 1.064 to -8.235). Likewise, by mutating C at pos. -4 to A (EYFP-6aa-AIIM>), the fusion protein remained in the cytosol (Figure 3.20, I). By contrast, the change of the C-terminal tripeptide IIM> to SKL> in both full-length AtIAN12 and the C-terminal domain construct caused re-direction of both constructs to different subcellular structures that coincided with CFP-labeled peroxisomes (Figure 3.20, J and K), consistent with a significant reduction in the prediction score for isoprenylation (from 1.064 to -8.565) and the well-known function of SKL> in directing proteins to peroxisomes.

Taken together, these results supported the idea that EYFP-AtIAN12 is first targeted to the ER for post-translational modification by isoprenylation at the C-terminal CIIM> motif and subsequently directed to small ER-derived subcellular vesicles. In onion epidermal cells and tobacco protoplasts these vesicles appear to be the final destination, while in tobacco mesophyll cells of intact leaves, EYFP-AtIAN12 was detected in peroxisomes (Figure 3.22, A-C), suggesting a third targeting step from the ER vesicles to peroxisomes.

Table 3.6: Summary of subcellular localization of AtIAN proteins

Construct name	Subcellular targeting	Figure
EYFP-IAN12 (CIIM>)	In tobacco leaves: peroxisomes In onions and tobacco protoplasts: organelle-like structures and interconnected punctate structures	Figure 3.22 Figure 3.20 & Figure 3.21
EYFP-IAN12 (ΔIIM>)	Cytosol	Figure 3.20
EYFP-IAN12 (GIIM>)	Cytosol	Figure 3.20
EYFP-IAN12 (CSKL>)	Peroxisomes	Figure 3.20
EYFP-6aa-CIIM>	Organelle-like structures	Figure 3.20 Figure 3.21
EYFP-6aa-CIIW>	Cytosol	Figure 3.20
EYFP-6aa-AIIM>	Cytosol	Figure 3.20
EYFP-6aa-CSKL>	Peroxisomes	Figure 3.20
EYFP-IAN11(CIIL>)	Organelle-like structures	Figure 3.23
EYFP-6aa-CIIL>	Organelle-like structures	Figure 3.23
EYFP-IAN8 (CSIL>)	Organelle-like structures	Figure 3.23
EYFP-6aa-CSIL>	Organelle-like structures	Figure 3.23

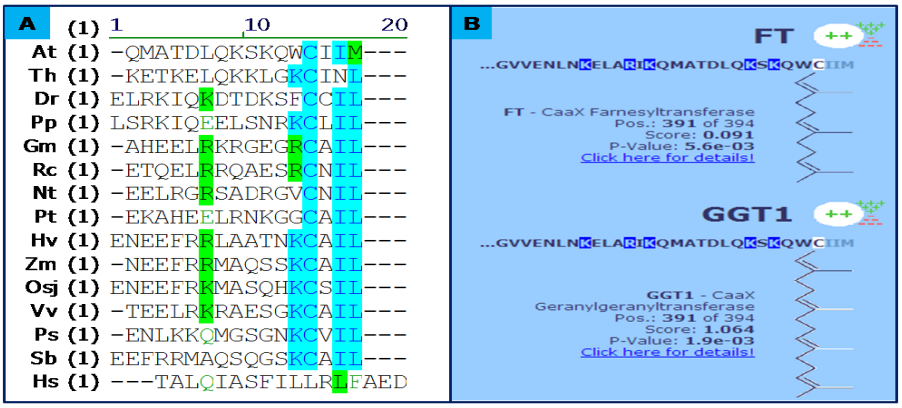


Figure 3.24: Conservation of the AtIAN12 isoprenylation motif (CIIM>)
A: Sequences of full-length protein AtIAN12 homologs were identified by BLAST search (NCBI) and aligned by AlignX program (Vector NTI, Invitrogen). The species abbreviations are as follows: At, *Arabidopsis thaliana*; Dr: *Danio rerio*; Gm: *Glycine max*; Hv: *Hordeum vulgare*; Hs: *Homo sapiens*; Nt, *Nicotiana tabacum*; Osj, *Oryza sativa japonica*; Pp, *Physcomitrella patens*; Ps, *Pisum sativum*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Sb: *Sorghum bicolor*; Tn: *Thellungiella halophila*; Vv, *Vitis vinifera*; Zm, *Zea mays*. B: prediction of AtIAN12 isoprenylation motif by http://mendel.imp.ac.at/sat/PrePS/ (Maurer-Stroh and Eisenhaber, 2005).

3.4.7 Generation of *AtIAN* overexpresser and amiRNA lines

To study the physiological function of AtIAN protein family members, T-DNA insertion lines were intended to be used but only a single line was available for *AtIAN11* (Sail_404_H08, see 2.2.1.5) in the wt Col-0 background. The T-DNA was inserted at the 300-UTRs of AtIAN11. Homozygous mutants (*ian11*) were isolated (Figure 3.25, A), similarly as described for *nhl* mutants (see 3.4.2). Two homozygous plants were identified and one of them was verified by making genomic PCR using either two gene-specific primers (LP and RP, for the wt allele) to confirm the absence of any wt allele and (LB1S and RP) to confirm the presence of the T-DNA insertion (Figure 3.25, B). In addition and similar to *NHL* genes (see 3.4.3), stable lines for *AtIAN* genes overexpression or knockdown by amiRNAs were created for all the three members of research focus (see Table 3.7).

Table 3.7: List of *AtIAN* overexpresser and amiRNA lines
Transformed seeds availability is indicated by (+), and the TF rate is indicated. T1 available lines number is indicated from the successfully genotyped plants. TF rates, expressed as 'percentage transformation', were calculated as [(#marker-resistant seedlings)/(total # seedlings tested)] x 100 (Clough and Bent, 1998).

No	Gene/ amiRNA	Plasmid/ promoter	TF Seeds	TF Rate (%)	T1 lines
1	*AtIAN12*	pBA002/35S	+	30	8
2	*AtIAN12*	pER10/Estradiol	+	0.25	8
3	*EYFP-AtIAN12*	pBA002/35S	+	40	11
4	*AtIAN8*	pER10/Estradiol	+	40	4
5	*AtIAN11*	pER10/Estradiol	+	0.14	2
6	*EYFP-AtIAN11*	Subcloned in pGEMT			
7	amiRNA (*AtIAN8*)	pER10/Estradiol	+	0.1	2
8	amiRNA (*AtIAN12*)	pER10/Estradiol	+	0.1	2
9	amiRNA (*AtIAN11+12*)	pBA002/35S	+	0.28	6

3.5.3 Proliferation of *Pst* DC3000 in *ian11* mutant

Bacterial growth of *Pst* DC3000 (avrRpt2) was monitored in *ian11* mutant plants. Together with *nhl* mutants (see 3.4.4) and under similar experimental conditions (n=2), the growth of *Pst* DC3000 (avrRpt2) was monitored in leaves of intact plants of *ian11* and compared with wt Col-0, *ndr1-1* (Century et al., 1995) and *pen2-2* (Lipka et al., 2005). The population of *Pst* DC3000 (avrRpt2) in wt Col-0 plants (2-4 days after inoculation) proliferated only 10 fold, while in *ndr1-1* and *pen2-2* mutants, bacteria proliferated 10,000 and 1000 fold, respectively. In *ian11* plants, bacteria proliferated around 1000 fold (Figure 3.25, D1). Disease symptoms appeared on the inoculated leaves for *ndr1-1* (2-4 d) after inoculation while did not show any symptoms on wt Col-0 (Figure 3.25, D2). Disease symptoms were less pronounced for *ian11*, while in *pen2-2*, leaf HR-like necrosis was most pronounced (Figure 3.25, D2). On the other hand in a single experiment, virulent *Pst* DC3000 proliferated similarly in *ian11* plants compared to the wt Col-0 (Figure 3.25, D2, single experiment). Taken together, *ian11* plants appeared to show less resistance to the avirulent strain than the wt, which indicates its possible important role in plant resistance.

3.5.4 Expression analysis of *AtIAN* genes

To study AtIAN8, AtIAN11 and AtIAN12 functions in plant responses, expression analysis by analysis of publicaly available microarray data and by real-time PCR were investigated. Briefly, the microarray available data indicate that *AtIAN8* is highly induced by a broad spectrum of biotic stresses, while *AtIAN11* and *AtIAN12* were also induced by different biotic stresses but to a lesser extent when compared with *AtIAN8* (data not shown). Preliminary data (n=1) for expression analysis of *AtIAN8*, *AtIAN11* and *AtIAN12* identified their pathogen-dependent mRNA accumulation. The real-time PCR conditions were similarly done as in *NHL* genes expression analysis (see 3.4.6). *AtIAN8*, *AtIAN11* and *AtIAN12* transcripts accumulated during the infection with the virulent *Pst* DC3000 (data not shown). *AtIAN8* only accumulated when the bacteria carried the avirulence gene *avrRpt2* (data not shown). Moreover, similar to the microarray data (data not shown), neither of *AtIAN8*, *AtIAN11* or *AtIAN12* transcripts accumulated after treatment of plants with flg22 which indicates that AtIAN proteins are not important in PTI (data not shown).

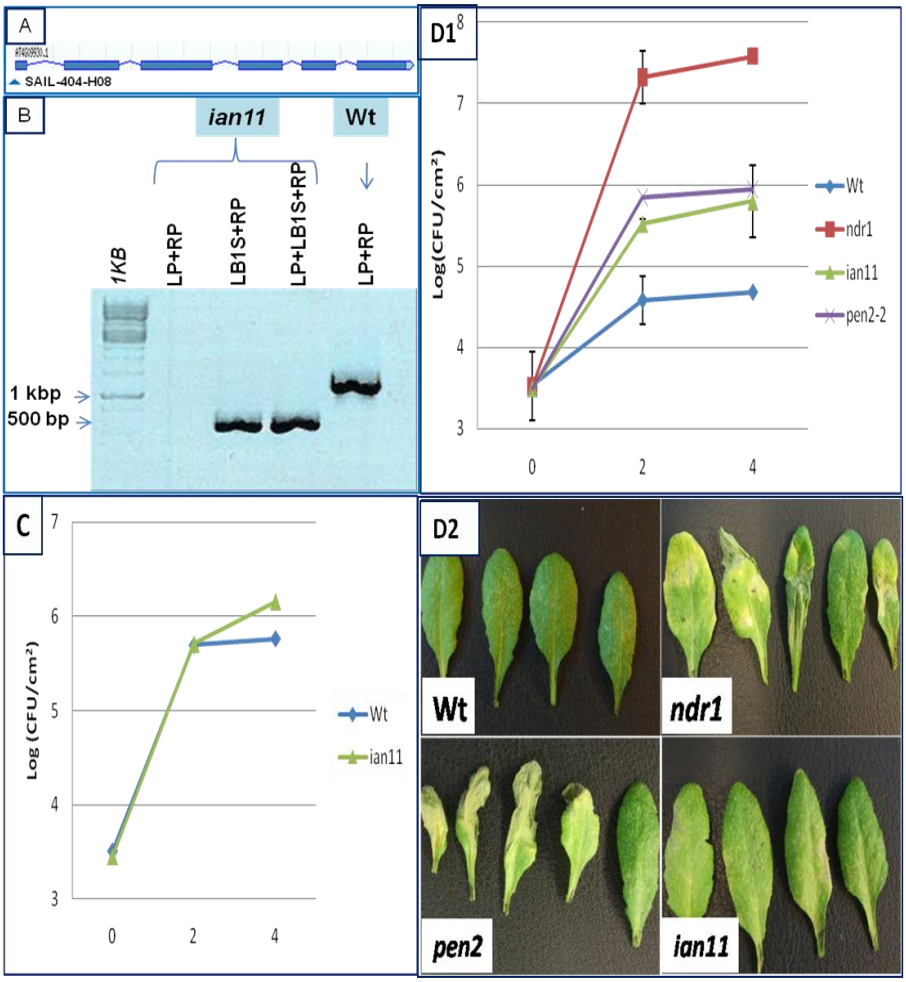


Figure 3.25: Identification of *ian11* and pathogen proliferation analysis
A: Diagram of the *Arabidopsis IAN11* gene. B: indicate the identified homozygous mutant by genomic PCR for *ian11* by different combination of primers as shown in the figure (wt band size = 1.2 kb, while the T-DNA specific band is ~500 bp). C: Pathogen proliferation analysis (virulent *Pst* DC3000) in wt Col-0 and *ian11* plants (n=1). D1: Pathogen proliferation analysis in wt Col-0, *ndr1-1*, *pen2-2*, *ian11* plants by avirulent *Pst* DC3000 (avrRpt2) (n=2, with higher SD, and needs further investigation. SD is calculated between the 2 experiments). D2: increased pathogen susceptibility. For C and D, mature soil-grown plants were infiltrated with low density *Pst* DC3000 (avrRpt2) or virulent *Pst* DC3000 solutions (concentration of 10^6 CFU/ml). The horizontal axis is in days. Bars show SD. For SD calculations see 2.2.4.1.

4. Discussion

Identification of the entire proteome of plant peroxisomes is crucial to understand all physiological functions of peroxisomes. The major focuses in the course of this study were experimental validation of novel PTS1 tripeptides and peroxisomal proteins identified by prediction algorithms, and understanding peroxisome functions in plant innate immunity. To investigate peroxisome functions in plant innate immunity, *Arabidopsis* proteins were screened for peroxisome-targeted PTS1 proteins with annotated functions related to plant defense against pathogens or stress responses. Several candidates were identified and their peroxisome targeting was validated by *in vivo* subcellular localization studies.

To get first insights into the molecular mechanisms of the validated peroxisomal defense proteins, two proteins (NHL4 and AtIAN12) were selected to initialize functional studies. Several members of the NHL and AtIAN protein families were found to carry predicted PTS1s and were subjected to *in vivo* subcellular localization targeting analyses as well. The functional studies, for NHL and AtIAN protein families, major purposes were to study their possible indirect involvement in metabolic peroxisome functions and to address if they are indeed important for plant innate immunity. Several steps were initiated to achieve these objectives 1) to generate homozygous T-DNA insertion lines, 2) to set up and apply immune-related assays on mutants, 3) to generate knockdown mutants using a siRNA approach, 4) to generate overexpression lines, and 5) to analyze their expression profiles under different biotic stress conditions by real-time PCR.

As part of a side-project, an investigation of the peroxisome function in H_2O_2 detoxification was initiated by *in vivo* subcellular analysis of several proteins (GR1, DHAR1, and GSTs) that previously were identified by experimental peroxisome proteomics. After the validation of GR1 and DHAR1 targeting to peroxisomes, genetic and molecular tools such as homozygous T-DNA insertion lines and recombinant proteins were generated to facilitate functional analyses in future studies.

4.1 Prediction models validation

The newly developed prediction models (PWM and RI, see 3. 1), predicted several novel plant PTS1 tripeptides and peroxisomal PTS1 proteins for *Arabidopsis*. Furthermore, the models were able to predict unknown low-abundance proteins. The models yielded high performance sensitivity and specificity values, allowing them to predict novel PTS1 tripeptides. Besides, the identification of several new PTS1 tripeptides from low-abundance proteins will allow searching for orthologous plant sequences, and most likely the recognition of further atypical PTS1s (see manuscript 1).

The accuracy of the prediction models was validated by extensive *in vivo* subcellular localization analyses. As part of the present dissertation, several predicted PTS1 tripeptides (SRV>, CKI>, STI>, AKM>, STI>, SPL>, PKI>, TRL>, LKL>, SGI>, and SEM>) were experimentally tested for functionality (manuscript 1, Table 1), where LCR>, LNL> were tested as cytosolic controls. The verification rate of predicted peroxisomal PTS1 tripeptides was high. For weak PTS1 tripeptides the sensitivity in detecting peroxisome targeting was improved by incubating the transformed tissue at low temperature for extended periods of time (from 24 h to 1 week expression time). All positive example sequences from the reliable data set that were tested experimentally were verified as functional PTS1 tripeptides (see 3.1.1). These data supported the high quality of the putatively orthologous sequences of this data set, and the accuracy of both prediction models (PWM and RI) on example sequences and the identification of several novel PTS1 tripeptides even including novel residues (manuscript 1, Table 1).

Furthermore, by applying the newly developed prediction models to the *Arabidopsis* genome (gene annotation of TAIR10), several proteins of unknown functions were predicted to be peroxisome-targeted by the PTS1 pathway. Some example proteins that were predicted and experimentally validated in the present dissertation included AC3, a Cys protease, S28FP, NUDT19, and pxPfkB (see manuscript 1, Figure 4, and Supplemental data set 2). These data supported the accuracy of the models to correctly predict low-abundance *Arabidopsis* PTS1 proteins (for more details see manuscript 1).

4.2 Detoxification-related proteins

The last two missing members of the peroxisomal ASC-GSH cycle, GR1 and DHAR1, had been identified at the molecular level by experimental proteome analysis of *Arabidopsis* leaf peroxisomes (see 1.1.1.2, (Reumann et al., 2007; Reumann et al., 2009)). Moreover, DHAR1, when fused C-terminally with EYFP, was detected in peroxisomes in intact tobacco leaves, but without determination of its targeting signal (Reumann et al., 2009). In the present study, validation of the peroxisomal localization of GR1 and identification of its C-terminal tripeptide (TNL>) as a novel functional PTS1 was accomplished (see manuscript 2). To investigate the location and nature of the PTS of DHAR1, DHAR1 was fused N-terminally with EYFP in order to investigate the possibility of the presence of unknown PTS1. This fusion protein, however, remained cytosolic, indicating that DHAR1 does not contain a PTS1 (Figure 3.1, A and J). On the other hand, the N-terminal domain was screened for the presence of any hidden PTS2-like structure. A PTS2-like structure ($RAx_{13}HL$) was found to be conserved in putative DHAR1 plant orthologs (Figure 3.2), and resembled the PTS2 nonapeptide motif $R[TMAV]x_5HL$ (Reumann, 2004) with the major difference that the four conserved residues are spaced by 13 rather than five residues. The N-terminal domain (46 aa) of DHAR1 was fused C-terminally with EYFP, and targeted to organelle-like structures (Figure 3.1, B). These data indicated that a PTS, and most likely a PTS2, is located in this N-terminal 46-aa domain of DHAR1. However, the mutation of the conserved R residue of ($\underline{R}Ax_{13}HL$) did not abolish organelle targeting (Figure 3.1, C), indicating that this peptide did not act as a PTS2. Thus, the identification of the PTS2 within this N-terminal 46-aa domain of DHAR1 remains elusive and requires further investigation that were beyond this side-project of this dissertation.

Similarly, five GSTs [see 1.1.1.2 (GSTT1, GSTU19, GSTU20, GSTF7, and GSTF10)], and HMGDH had been identified at the molecular level by experimental proteome analysis of *Arabidopsis* leaf peroxisomes (Reumann et al., 2007; Reumann et al., 2009). Four GSTs and HMGDH lacked recognizable PTS-like peptides. In this study, the identified proteins were fused N-terminally with EYFP. The reporter fusions were investigated by *in vivo* subcellular localization analyses,

but remained in the cytosol (Figure 3.1, E-I and K-L). EYFP-GSTF10 was also detected mostly in the cytosol but also in organelle-like structures in a limited number of transformants whose identity could not be further investigated due to weak organelle targeting efficiency (Figure 3.1, G). These data indicate the absence of PTS1s in the investigated proteins. The peroxisomal identity of the GSTs remains to be verified by C-terminal reporter protein fusion studies. Alternatively, one could search for interaction partners that contain PTSs and could transfer these proteins to peroxisomes by piggy-backing (Kaur et al., 2009).

The validation of GR1 and DHAR1 targeting to peroxisomes prompted us to initiate functional analyses of the ASC-GSH cycle. Heterologous overexpression of DHAR1 and GR1 in *E. coli* was largely accomplished. GR1 and DHAR1 were successfully expressed as soluble MBP-tagged proteins, and DHAR1 was also produced as a soluble His_6-tagged protein (Figure 3.4). The availability of both enzymes *in vitro* will allow analysis of their function by determination of their kinetic properties. Homozygous *gr1* and *dhar1* mutants were also isolated from T-DNA insertion lines. As a quick screen for deficiencies in peroxisome metabolic functions, photorespiration and β-oxidation assays were applied to *gr1* and *dhar1* plants. It was found that these metabolic functions of peroxisomes were not affected to major extent in both mutant plants as compared to the wt (Figure 3.3). Furthermore, several *Arabidopsis* leaf peroxisome fractions were isolated from mature leaves by the two-density gradient approach (Reumann et al., 2007). The isolated leaf peroxisomes will be used in future studies for downstream biochemical analyses of the ASC-GSH cycle in order to determine the activities of GR1 and DHAR1.

4.3 Peroxisome defense-related proteins

Apart from maintaining redox homeostasis under different stress conditions, much information about peroxisome functions regarding defense against pathogen is unknown. Indeed, peroxisome functions in plant innate immunity were reported only recently (see 1.1.1.3). To be able to understand these functions in greater details, it is important to identify further peroxisomal defense-related proteins other than SGT and PEN2 (see 1.1.1.3). *Arabidopsis* proteins were screened using the newly developed prediction algorithms (Lingner et al., 2011) for peroxisome-targeted PTS1 proteins with annotated functions related to

plant defense against pathogens or stress responses. Several unknown candidate proteins studied as defense-related but which had not been linked before with peroxisomes, were predicted as peroxisomal proteins (Table 3.2). Moreover, in latest and relatively high quality proteome studies of mature leaf peroxisomes, two stress-related proteins were also identified, AtMIF1 and AtOZI1 (Reumann et al., 2007; Reumann et al., 2009). *In vivo* validation of the predicted defense-related proteins was further investigated for their full-length proteins and/or for their PTDs by their N-terminal fusions with EYFP. Several defense-related proteins were validated to be peroxisomal such as AtMIF1, AtSurE, AtLIMDP and AtCAD7. Two other defense-related proteins also gave strong indications to be peroxisomal such as AtDRP and AtCDR1. The details for each of the identified proteins will be addressed for each one separately.

AtMIF1 (a homolog of human MIF that is important immune-regulator molecule in human) had been identified by proteome analysis (Reumann et al., 2007; Reumann et al., 2009) and was validated by subcellular localization analysis in peroxisomes in the present study (Figure 3.5, A and B). The protein has the prototypical PTS1 tripeptide (SKL>). In parallel to this study, similar data on peroxisome targeting validation for AtMIF1 were published. The authors established a new method for transient expression, referred to as fast *Agrobacterium*-mediated seedling transformation (FAST) using AtMIF1 as an example protein for peroxisome targeting (Li et al., 2009), consistent with the peroxisome targeting data for AtMIF1 in onions and tobacco protoplasts obtained in the present study. AtMIF1 might have important roles in plant resistance responses towards biotic stresses, because *AtMIF1* is highly induced by a broad spectrum of biotic and abiotic stress conditions (Figure 3.7 and Table 3.1). In *Arabidopsis*, there are two other homologs of AtMIF1 that were referred to as AtMIF2 and AtMIF3. Both proteins have less pronounced stress-related induction patterns and were detected in a chloroplast proteome study (Zybailov et al., 2008). Despite its PTS1-like tripeptide (ATL>), EYFP-AtMIF2 remained in the cytosol. In order to initiate functional studies, one T-DNA insertion line for *AtMIF1* was obtained but several trials to isolate homozygous mutants failed. Other trials need to be further pursued in the near future. It will be very important to obtain a conditional knockout and/or knockdown mutant for AtMIF1 to be able to test different pathogen assays on the loss-of-function mutant to investigate

the protein's postulated role in plant innate immunity. Notably, human MIF is an immune cytokine that is released from the peripheral immune cells and pituitary gland, and acts as a signal in immune regulation and has a central role in the development of innate and adaptive immune systems (Golubkov et al., 2006). It was postulated previously that peroxisomes are involved in preinvasion defense against fungi via PEN2 and PEN3 (see 1.1.1.3 and Figure 1.3, D). Moreover, because peroxisomes accumulate at infection sites, they might release different toxic and signaling molecules similar to immune vesicles (see 1.1.1.3). By combination of the above mentioned postulations and facts, AtMIF1 might mediate signaling in response to pathogen infection. AtCDR1 is another protein that is predicted to be in peroxisomes, and reported to be involved in SAR signaling in *Arabidopsis*, where it was found to accumulate in intercellular fluid in response to pathogen attack (see 1.2.2.3, (Xia et al., 2004)).

AtSurE is a homolog of SurE which has activities as a nucleotidase and an exopolyphosphatase and is thought to be in involved in stress response in *E. coli* (Proudfoot et al., 2004). AtSurE was identified by the new PTS1 prediction algorithms (SSL>, (Lingner et al., 2011)). AtSurE is a constitutive protein but was also detected by microarray analysis to be induced by biotic and abiotic stresses (Figure 3.7 and Table 3.1). The EYFP-AtSurE fusion protein was targeted to non-peroxisomal organelle-like structures in both onion epidermal cells (Figure 3.5, C), and tobacco protoplasts 24 h P.T. (Figure 3.5, G). In both expression systems the reporter fusion was also detected in aggregate-like structures in a considerable number of transformants (Figure 3.5, D and H). The pattern of these aggregate-like structures was not previously observed during the present study and indicates that a large number of small punctate structures must have aggregated, or the fusion protein was accumulated somehow intensively in unknown structures and failed to be exported. In contrast, EYFP-AtSurE was detected in peroxisomes 48 h P.T. in tobacco protoplasts (Figure 3.5, F). The identification of AtSurE as peroxisomal protein was further supported after the identification of the C-terminal domain of AtSurE PTD as a functional PTS1 domain (Figure 3.5, E). These data indicate an indirect transport of AtSurE to peroxisomes through an intermediate step that remains elusive and was beyond the scope of the present study. In order to initiate functional studies, one AtSurE T-DNA insertion line was obtained and homozygous mutants were successfully

isolated. A first mutant screen indicated that basic metabolic functions of peroxisomes were not severely affected in *atsure* plants (data not shown).

AtDRP and AtLIMDP belong to the R protein classes TIR-NBS-LRR and CC-NBS-LRR (see 1.2.1.2), respectively, members of which are implicated in signal transduction leading to ETI (see 1.2.1.2). Both proteins were identified by PTS1 prediction algorithms to contain PTS1 tripeptides (CRL> and SKL>, respectively). Both *AtDRP* and *AtLIMDP* appear to be induced upon different biotic and abiotic stresses based on their expression patterns deduced from publicly available microarray experiments (Figure 3.7 and Table 3.1). Recently AtLIMDP, also referred to as CHS3, was reported to be important in defense response and chilling tolerance. In this study, *chs3-1 Arabidopsis* mutant plants showed arrested growth, chlorosis, and exhibited constitutively activated defense responses when grown at 16^{o}C, which were alleviated at 22^{o}C (Yang et al., 2010), but the protein targeting was not investigated in the mentioned study. In the present study, EYFP-AtLIMDP (C-terminal 472 aa) was targeted to peroxisomes (Figure 3.6, E). For the full-length protein (1613 aa), a cloning strategy has been recently developed and will be pursued in the near future. On the other hand, EYFP-AtDRP was targeted to peroxisomes in onion epidermal cells (Figure 3.6, C). It is noteworthy mentioning that during transient expression of *EYFP-AtDRP* in three different experiments in onion epidermal cells, the detection level of both fluorophores (EYFP, or CFP for the peroxisomal marker) was very low for unknown reasons. Long acquisition times were needed for capturing images (Figure 3.6, C). For this reason, and to further confirm its peroxisome localization, this construct is recommended to be expressed in other expression systems. The C-terminal domain of AtDRP PTD was validated to be functional PTS1 domain (Figure 3.6, D), and hence the CRL> tripeptide as a functional PTS1 further supported the prediction. *AtDRP* has two transcriptional variants as proposed by TAIR10 (Figure 3.26, A), where variant number 2 is the one represented in the present study because of its C-terminal tripeptide CRL>. The peroxisomal verification of these two R proteins will shift the thoughts about possible modes of action of peroxisomes in plant innate immunity. Presently, the only proposed role of peroxisomes is that they contain enzymes that synthesize toxic molecules that are released into the apoplast (see 1.1.1.3 and Figure 1.3, D). The identification of these two R protein homologs is strongly

suggesting that peroxisomes could be also involved in pathogen recognition and further downstream signal transduction pathways. R proteins are usually PM-associated and are implicated in signal transduction (see 1.2.1.2). In order to initiate functional studies, one T-DNA insertion line for *AtLIMDP* was obtained and homozygous mutants were successfully isolated. A first mutant screen indicated that basic metabolic functions of peroxisomes were not severely affected in *atlimdp* plants (data not shown).

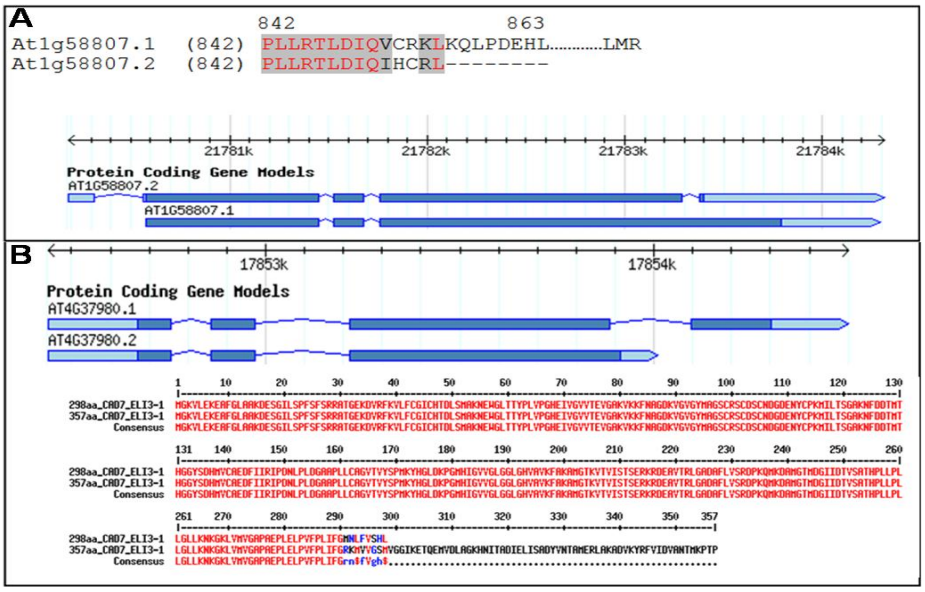


Figure 3.26: Transcriptional variants of *AtDRP* and *AtCAD7*
A: AtDRP C-terminal alignment and TAIR proposed protein coding gene models indicate the difference between the two variants. B: AtCAD7 alignment and TAIR proposed protein coding gene models demonstrate the difference between the two variants. The AtCAD7 variant 1 is 357 aa long, while variant 2 is 298 aa long.

Another two proteins were identified, AtCAD7 and AtCDR1, by PTS1 prediction algorithms. Interestingly, the reporter fusion of AtCAD7 was detected in peroxisomes in the present study (Figure 3.6, A), also referred to as ELI-3, which has been implicated in defense response. ELI-3 was identified in parsley after treatment by a heat-released elicitor from the fungus *Phytophthora rnegasperma*, f. sp. *Glycinea*, and in *Arabidopsis* after treatment by fungal elicitor. Additional

evidence for an important role of the ELI-3 in plant disease resistance came from genetic studies demonstrating that *ELI-3* expression was dependent on the presence of *RPMI* (see 1.2.1.2) in *A. thaliana* (Somssich et al., 1989; Debener et al., 1991; Kiedrowski et al., 1992; Trezzini et al., 1993). Furthermore, the C-terminal domain of AtCAD7 PTD was also validated to be functional PTS1 domain, as the fusion protein EYFP-PTD was detected in peroxisomes (Figure 3.6, B), and hence SHL> is a functional PTS1. SHL> has been characterized previously as a functional but weak PTS1 whose peroxisome targeting capability depends on auxiliary targeting enhancing elements immediately upstream of the tripeptide (Ma and Reumann, 2008). *AtCAD7* has two transcriptional variants as proposed by TAIR (Figure 3.26, B), where the short variant number 2 (298 aa) is the one investigated in this study. The fusion protein EYFP-AtCDR1 remained in the cytosol and was not detected in peroxisomes (Figure 3.6, G) in onion epidermal cells. However, the C-terminal domain of AtCDR1 PTD was found to be functional PTS1 domain, as the fusion protein EYFP-PTD was detected in peroxisomes (Figure 3.6, H), and hence AKM> was found be functional PTS1. AtCDR1 is an aspartic proteinase involved in disease resistance. AtCDR1 was also reported to be involved in SAR signaling in *Arabidopsis*, where it was found to accumulate in intercellular fluid in response to pathogen attack. Moreover, AtCDR1 was implicated in SAR signaling by generating small mobile signals (see 1.2.2.3 (Xia et al., 2004; Simöes et al., 2007)). Why did the full-length AtCDR1 remain in the cytosol? This question shall be addressed in the upcoming studies, especially with the confirmation that it has a functional PTS1 tripeptide. The full-length protein shall be further analyzed in tobacco protoplasts, and most preferably in intact tobacco leaves.

4.4 NHL protein family investigations

4.4.1 PTS1 prediction and sequence analysis of NHL proteins

Using PTS1 predictions algorithms, seven NHL protein (see 1.2.3.1) family members from *Arabidopsis* were predicted to carry a potential PTS1 (Table 3.3). In this study, the subcellular localization of all predicted members was investigated, with the focus on three NHL candidate proteins with highest PTS1 protein prediction scores, namely NHL4, NHL6, and NHL25, all of which belong to one phylogenetic

clade among 45 identified *Arabidopsis* NHL homologs (Figure 3.8). By protein multiple sequence alignment analysis of PTS1 predictable proteins with selected NHLs and tobacco HIN1 (Figure 3.9), the three studied members were confirmed to be similar to *Arabidopsis* NDR1, tobacco HIN1 and NHL members that were implicated in disease resistance, for example NHL2, NHL3 and NHL10 (see 1.2.1.2). The three candidates have three conserved motifs, and the WHy domain, all of which are conserved in NHL family and are found in tobacco HIN1. The WHy domain is also found in the LEA-14 family [expressed during embryogenesis and in plant responses to desiccation (extreme drying)] which suggests a shared mechanism in plant response during HR and desiccation stresses (Ciccarelli and Bork, 2005).

Additionally, NHL4, NHL6, and NHL25 were predicted to be membrane proteins similar to NDR1 and NHL3 (see 1.2.1.2). The computational analysis using transmembrane prediction programs, DAS-TM filter [http://mendel.imp.univie.ac.at/sat/DAS/DAS.html, (Cserzo et al., 2002)] indicated a single putative transmembrane helix domain (NHL4: 54-78, NHL6: 67-90, NHL25: 62-87). This domain was also predicted as an uncleavable signal anchor by SignalP-3.0 [www.cbs.dtu.dk/services/SignalP, (Bendtsen et al., 2004)]. NDR1 was reported to be localized to the PM via a C-terminal GPI-anchor (see 1.2.1.2), and is suggested to be located in the outer surface of the PM to act in a pathogen signal transduction cascade (Coppinger et al., 2004). A GPI anchor was not predicted for NHL4, NHL6, and NHL25. This is supporting the localization data that no PM-association was detected. Identification of protein sequence properties for the studied candidates implies that they are definitely important in pathogen response, and the possibility to be organelle membrane associated. Because of time limitations and the intensive studies on subcellular targeting analyses for NHL proteins, more detailed investigation of the possible organelle membrane association was beyond this study. The major focus was to identify the in *vivo* subcellular localization of the predicted candidates, and further initiation of functional studies to understand their function in plant innate immunity.

4.4.2 *In vivo* subcellular localization of NHL proteins

Identification of the subcellular localization of NHL proteins will assist in the elucidation of their biochemical function and mode of action in

plant responses to biotic stresses. NHL4 has a high PTS1 prediction score (AKL>, see Table 3.3), but EYFP-NHL4 unexpectedly was targeted to non-peroxisomal organelle-like structures in onion epidermal cells even after extended cold incubation (Figure 3.10, A). On the other hand, and consistent with the prediction, EYFP-NHL4 was detected in peroxisomes in tobacco protoplasts (Figure 3.11). NHL4 was thereby the first NHL member to be associated with peroxisomes. The identification of this protein in peroxisomes is giving more indications of the peroxisome functions in plant innate immunity.

Notably, EYFP-NHL4 targeting to peroxisomes was indirect, and the fusion protein was found to change its localization to peroxisomes in a time-dependent manner of expression in tobacco protoplasts. Based on the experimental data (Figure 3.11), the EYFP-NHL4 fusion protein is suggested to target peroxisomes through three successive steps, 1) first to be targeted to non-peroxisomal organelle-like structures (Figure 3.11, A), 2) the organelle-like structures dock to the peroxisome surface (Figure 3.11, B-D) and 3) the organelle-like structures fuse with the peroxisomal membrane and release their matrix content into peroxisomes (Figure 3.11, D and E). To better explain the hypothetical import pathway of NHL4 into peroxisomes: NHL4 might be inserted into specific ER domains, that pre-peroxisomes (organelle-like structures) bud off from the ER and that these pre-peroxisomes merge with mature peroxisomes similar to PMP1s (see 1.1.2, Figure 1.4). Notably, in onion epidermal cells these putative pre-peroxisomes do not appear to merge with mature peroxisomes. In the present study, AtSurE was also detected in onion epidermal cells in putative pre-peroxisomes or vesicles, which were identified in tobacco protoplasts to be in peroxisomes in a time-dependent manner. It is worth mentioning that research studies from oilseed rape (*Brassica napus*) identified two NHL homolgs, namely BnNHL18 (A and B), and their subcellular localization was also reported to change from the ER to an unknown destination according to their studies, when expressed in protoplasts subjected to stress treatments [sodium chloride, H_2O_2, JA and SA, (Lee et al., 2006)]. These data indicate that these protein family members could change their localization according to stress conditions.

Similarly, NHL6 and NHL25 reporter fusion proteins were targeted to non-peroxisomal organelle-like structures in onion epidermal cells

(Figure 3.10, B and C). Interestingly, EYFP-NHL25 was targeted to the same organelle-like structures which were labeled with CFP-NHL4 in onion epidermal cells (Figure 3.12, B). Moreover, preliminary data for EYFP-NHL6 show, in a limited number of transformed onion epidermal cells, coincidence with CFP-NHL4 within the same structures (Figure 3.12, A). These data indicate that both NHL6 and NHL25 target peroxisomes as well and share a common import pathway with NHL4. Another indication for peroxisomal targeting of both NHL6 and NHL25 was the detection of their EYFP-PTDs in peroxisomes (Figure 3.12, C and D) and establishing their PTS1s as a functional tripeptides (LRL>, NHL6 and FRL>, NHL25), consistent with PTS1 predictions. Finally, other predictable PTS1 NHL proteins (NHL39, NHL39H1, NHL13H1, and NHLx), all of which were investigated by the *in vivo* subcellular localization in onion epidermal cells. The four fusion proteins targeted to unidentified organelle-like structures in onions (Figure 3.13). These data indicate that NHL family members are targeted to peroxisomes and are likely to give more insights into peroxisome functions in plant innate immunity in the near future. Earlier studies reported targeting of NHL3 to PM (Varet et al., 2003), and NHL2 and NHL10 to chloroplast (Zheng et al., 2004). Overall, these family members appear to have different mode of actions according to their different subcellular localization.

4.4.3 Generation of transgenic lines for reverse genetic analyses

To initiate molecular analysis for NHL4, NHL6 and NHL25, T-DNA insertion lines were obtained and homozygous mutants were successfully isolated (*nhl4*, *nhl6*, and *nhl25*). Moreover amiRNA stable lines were generated to be used for future analyses. Previously, loss-of-function mutants of NHLs were not adequately studied, except for NDR1. *ndr1-1* (fast-neutron-generated mutant, see 1.2.3.1) is reported to be more susceptible to several avirulent *Pst* DC3000 strains containing one avirulence gene [*avrB*, *avrRpm1*, *avrRpt2*, or *avrPph3*, (Century et al., 1995; Century et al., 1997)]. Moreover, the generation of overexpresser stable lines was largely accomplished for *NHL4* and *NHL25*. The overexpresser lines for each gene were produced by both a constitutive (35S) and an inducible (estradiol) promoter. The overexpresser lines will be used to investigate gene overexpression effects on plant response to pathogen infection. Several members from this family were identified to confer resistance (Varet et al., 2003;

Coppinger et al., 2004), and affect plant responses towards pathogen infection (see 1.2.3.1).

4.4.4 Functional studies on NHL protein family

To investigate whether the NHL proteins of interest indirectly participate in fatty acid β-oxidation and IBA-to-auxin conversion of peroxisomes, the available homozygous mutants (*nhl4*, *nhl6* and *nhl25*) were subjected to sucrose dependence, and auxin assays. *pex14* plants showed growth inhibition in the absence of sucrose and also were insensitive to IBA, both of which is consistent with being as a positive control for these assays. In this mutant peroxisome functions were altered due to the absence of PEX14. *nhl* mutants were similar to wt in the results obtained from both assays. The data led to the conclusion that the absence of NHL4, NHL6 and NHL25 did not affect the two metabolic functions of peroxisomes (Figure 3.17).

To investigate the functions of the selected genes in plant response to pathogens, the corresponding infection assays were first established in the group, as learned previously in the He group (PRL, MSU, USA). One crucial characteristic of resistant plants is the ability to restrict *in planta* growth of avirulent bacteria. To be able to monitor plant resistance to pathogens, the bacterial proliferation assay was established. *Pst* DC3000 virulent and avirulent strains (e.g., carrying *avrRpt2*) were obtained and used for analysis of *in planta* bacterial proliferation (Figure 3.15, A). Using this assay, the knockout mutants (*nhl4*, *nhl6*, and *nhl25*) were subjected for checking their possible susceptibility to pathogen infection. Preliminary results indicated that virulent bacteria proliferated at similar rate in *nhl* mutants and wt plants upon leaf infiltration (Figure 3.15, B). In contrast, the avirulent (avrRpt2) strain proliferated differently in *nhl* mutants indicating elevated pathogen susceptibility (Figure 3.15, C1). The rates of bacterial proliferation were increased for the positive controls *ndr1-1* and *npr1-1* mutants (Figure 3.15, C1) as reported previously (Cao et al., 1994; Century et al., 1997). Because of the presence of the effector protein avrRpt2 in the avirulent strain, the plants should be more resistant, and the bacteria should grow slowly as indicated in wt Col-0 plants, because of the induction of ETI. In contrast, the avirulent bacteria grew at high rate in both positive controls and *nhl* mutants (Figure 3.15, C1), which indicates that the plant immune system was

not up-regulated, and the plants were more susceptible to bacterial infection. These results suggest that NHL4, NHL6 and NHL25 are involved in ETI (see 1.2.1.2). However, it needs to be pointed out that the proliferation assays remained preliminary due to an insufficient number of only two biological replicates in independent experiments. Several trials were done to generate 3rd repetition, but these were hindered by technical difficulties with plant growth chamber facility. The plants generated in the last trials sufferd and were delayed in growth because of the growth conditions. It is crucial to obtain very healthy and immune plants to be able to have a successful assay. However, some healthy plants were selected from these conditions and subjected for the assays; but unfortunately, varied proliferation rates of bacterial growth in the wt Col-0 was obtained and halted the interpretation of the complete assay.

PAMP-induced callose deposition in cotyledons of hydroponically grown *Arabidopsis* seedlings was carried out in duplicate experiments in another lab at Hu group (PRL, MSU, USA). The time scale was too limited to establish the system and reproduce the same data in the Reumann lab. It is important to mention that a recent study on this assay indicated that callose is a multifaceted defense response and is controlled by distinct signaling pathways, depending on the growth conditions (Luna et al., 2011). In the present study, when applying flg22 to plant seedlings, wt Col-0 plants showed a high average number of callose depositions indicative of functional PTI response, which were absent in wt WS-0 (Figure 3.16), consistent with this ecotype representing a negative control for this assay because lacking the FLS2 receptor that is responsible for flg22 recognition. Furthermore, analysis of callose depositions for *nhl4*, *nhl6*, *nhl25* and *pen2-1* mutants after the same treatment indicated varied numbers of callose depositions between plants. On average, five representative leaves from five independent plants were used for calculations. These numbers indicated that the mutants were able to induce callose formation but at significantly lower frequency than wt Col-0. To conclude, these data shall be reproduced, and *pen2-1* mutant shall be replaced by another negative control, because previously it was shown that callose depositions are completely absent (Clay et al., 2009; Luna et al., 2011). Another possibility to support these results is to investigate callose depositions after infiltrating 4-week-old leaves by the PAMP (Galletti et al., 2008), and further compare with the results from hydroponic

grown seedlings. This will help in getting rid of the possible invariability which is produced by hydroponic growing seedlings.

4.4.5 Expression analysis of *NHL* genes

NHL proteins are proposed to be possible mediators in plant resistance against pathogen infections. The mechanism of their action, however, is not well-known, for instance, if they are required as R function or involved in defense gene activations (Dormann et al., 2000). Several family members were induced during pathogen infections and signaling molecule treatment (see 1.2.3.1). To study possible functions of NHL4, NHL6 and NHL25 in plant responses, expression analysis of available microarray data and real-time PCR were carried out. Briefly, the microarray data (Figure 3.18) indicated that *NHL6* is highly induced by a broad spectrum of biotic stresses. *NHL4* similarly is induced but less pronounced than *NHL6*. In the present study, preliminary mRNA transcript quantification using real-time PCR identified the induction of the three genes by virulent *Pst* DC3000 (Figure 3.19, A) and SA (data not shown, M.Sc. thesis, (Mwaanga, 2011)) while specifically *NHL6* and *NHL25* were induced by avirulent (Figure 3.19, B) *Pst* DC3000 (avrRpt2). Interestingly, *NHL6* specifically was induced after the treatment by flg22 (Figure 3.19, C). To dissect if NHL6 is induced downstream the recognition of FLS2 for flg22, another experiment was applied on *fls2* mutant plants. In contrast to wt plants, *NHL6* transcripts did not accumulate in *fls2* mutant plants after the treatment by flg22 (Figure 3.19, D). Taken together, these results indicate that *NHL6* is induced through FLS2 recognition of pathogens as part of PTI (see 1.2.1.1).

NPR1 is induced through SA signal transduction. *Arabidopsis NPR1* is a key regulator of SAR (see 1.2.2.3), and it is highly expressed upon infection, in turn inducing expression of a battery of downstream *PR* genes (see 1.2.2.3) through binding to TGA2 transcription factor in the nucleus (Cao et al., 1998; Subramaniam et al., 2001). Because *NHL6* is already found to be induced by SA, it worth investigating if *NHL6* is induced downstream or upstream of NPR1. *NHL6* transcripts were monitored in *npr1-1* mutants after flg22 treatment, and found to be accumulated similarly to wt plants (Figure 3.19, D). These preliminary data indicate that *NHL6* induction is NPR1-independent and that NHL6 appears to be located upstream of NPR1 in the SA induction cascade.

Even though, these data are preliminary, the flg22 treatment of wt plants similarly gave the same induction fold increase for *NHL6* in two different experiments, and is strongly supporting the expression analyses in this study.

4.5 AtIAN protein family investigation

The IAN protein family has been largely studied in humans and has important role in immune responses (see 1.2.3.2). In plants, AtIAN8/AIG1 has previously been identified and partially been studied upon pathogen infection (Reuber and Ausubel, 1996). This study initiates the first molecular analyses of plant IAN proteins after the identification of one AtIAN8 homolog (AtIAN12) by *Arabidopsis* leaf peroxisome proteomics (Reumann, unpub. data). *AtIAN*s are mostly clustered on chromosomes I, II, and IV. This study focuses on *AtIAN8*, which is located on chromosome I, and one phylogenetic clade comprising *AtIAN11* and *AtIAN12*, which are located on chromosome IV (Liu et al., 2008).

4.5.1 *In vivo* subcellular localization of *AtIAN* proteins

In this study, subcellular localization studies validated AtIAN12 targeting to peroxisomes after transient expression in intact tobacco leaves using the *Agrobacterium*-mediated transformation method. The EYFP-AtIAN12 was detected in organelle-like structures, which coincided with CFP-labeled peroxisomes in intact tobacco leaves (Figure 3.22). However, the reporter fusion protein was targeted to organelle-like structures and interconnected organelle-like structures in onion epidermal cells, and tobacco protoplasts (Figure 3.20 and Figure 3.21). But peroxisomal validation was not possible in both of these expression systems. Similarly, AtIAN8 and AtIAN11 were targeted to organelle-like structures in onion epidermal cells and tobacco protoplasts that could not be shown to be identical with peroxisomes (Figure 3.23). The transient expression of both AtIAN8 and AtIAN11 in intact tobacco leaf cells was delayed because of the unavailability of their EYFP-cDNAs in a binary vector. Additionally, the EYFP-PTDs for the three selected AtIANs were mostly targeted to non-peroxisomal organelle-like structures (Figure 3.23). Taken together, these data mostly could not validate peroxisome targeting, but indicated that the targeting signal is located in the C-terminal decapeptides.

Initially, the C-terminal IIM> of AtIAN12 was suggested to be an atypical PTS1 based on its similarity to the weak PTS1 (AKM>). Methionine is an abundant and "strong" residue at pos. -1, isoleucine has both been shown to represent an allowed PTS1 tripeptide residue at pos. -3, and pos. -2 is presently considered the most flexible residue with 15 allowed amino acid residues, even though presently excluding isoleucine [(Lingner et al., 2011), Figure 1, B]. In addition, the three AtIANs were shown to carry predicted isoprenylation motifs in the C-terminal domain CaaX>; (AtIAN12: CIIM>, AtIAN11: CIIL>, and AtIAN8: CSIL>), offering the possibility that EYFP-PTDs of the three AtIANs might be targeted to and anchored in the membrane of endomembrane vesicles via attachment of an isoprenyl moiety.This prediction analysis was tested by Prenylation Prediction Suite, PrePS (http://mendel.imp.ac.at/sat/PrePS/).

The peroxisomal AtIAN12 was further validated for the presence of isoprenylation motif. Notably, the prediction motif CaaX> was found to be conserved in its plant orthologs (Figure 3.24). Several point mutations and deletions were introduced into *AtIAN12* full-length cDNA and PTD to generate EYFP-AtIAN12 (ΔIIM> and GIIM>) and PTD (AIIM>, and CIIW>), in order to validate functionality of the isoprenylation motif. As predicted all the applied mutations made the fusion proteins to remain cytosolic and strongly indicated CIIM> as a functional isoprenylation motif (Figure 3.20). The protein isoprenylation pathway in plant cells includes an intermediate step at the ER for the cleavage of the tripeptide aaX and C-terminal methylation (Crowell, 2000; Galichet and Gruissem, 2003). Preliminary data indicated that EYFP-AtIAN12 was partially detected in the ER upon transient expression in tobacco protoplasts (Figure 3.21, D and E), while this was not the case upon transient expression in tobacco leaves (Figure 3.22, E). The absence of AtIAN12 in the ER in tobacco leaves could indicate that the isoprenylation modification was active and that peroxisome targeting was accelerated. In contrast, in tobacco protoplasts, protein isoprenylation might be less active, thereby halting EYFP-AtIAN12 in the ER and ER-export vesicles. Taken together, these results indicate that EYFP-AtIAN12 is isoprenylated, and targeted to the ER for IIM> cleavage and C-methylation, and subsequently directed to small ER-derived subcellular vesicles. In onion epidermal cells and tobacco protoplasts, these organelle-like

structures appear to be the final destination, while in tobacco mesophyll cells of intact leaves; EYFP-AtIAN12 was detected in peroxisomes. Furthermore, AtIAN12 might be exported to peroxisomes under special conditions such as during upregulation of immune systems because of pathogen infection. Consistent with this hypothesis, *Agrobacterium* could led to upregulation of the immune system of the plant and gave the conditions needed for AtIAN12 targeting to peroxisomes from ER in *Agrobacterium*-mediated transformation in intact tobacco leaves. This hypothesis can be tested on the available EYFP-AtIAN12 stable lines, where if AtIAN12 will not be detected in peroxisomes, the plants could be infected by *Agrobacteria* and to investigate the possible peroxisome targeting of EYFP-AtIAN12 afterwards.

Only little information is available on isoprenylation of peroxisomal proteins for all three kingdoms to date. Yeast PEX19p is farnesylated and seems essential for proper matrix protein import into peroxisomes through possible induction of a conformational change in PEX19p, and hence affecting on PMPs (Rucktaschel et al., 2009). Moreover, Rho family members (G-proteins) that are known to play a role in actin reorganization and membrane dynamics also contain the isoprenylation consensus sequence and are reported to have a role in plant response towards infection (Goritschnig et al., 2008). Interestingly, one small GTPase (Rho1p, YPR165W) localizes to peroxisomes through interaction with the PMP PEX25, and regulates the assembly state of actin on the peroxisome membrane (Marelli et al., 2004). However has not been reported previously that Rho1p includes a predicted isoprenylation motif (CVLL>, score 3.065). If Rho1p indeed uses the isoprenylation pathway for peroxisome targeting, this would be another link on the importance of isoprenylation pathway on sorting proteins to peroxisomes and possibly to interact with PMPs. Noticeably, protein farnesylation was reported to play a role in plant innate immunity, because the mutant *era1* (enhanced response to ABA 1), which has a defect in the enzyme farnesyltransferase, is more susceptible toward virulent bacterial and oomycete pathogens (Goritschnig et al., 2008). Finally, another possible experimental evidence for AtIAN12 farnesylation confirmation could be by studying the subcellular localization of AtIAN12 in the mutant *era1*, to determine if the farnesyltransferase absence could affect on AtIAN12 targeting and make it to remain in the cytosol.

4.5.2 Generation of transgenic lines for reverse genetic analyses

To initiate molecular analyses for AtIAN proteins, T-DNA lines for AtIAN11 were only available, and homozygous plants were successfully isolated (*ian11*). The generation of amiRNA stable lines for AtIAN12, AtIAN11 and AtIAN8 was largely accomplished. Also, one amiRNA stable line was generated to silence both *AtIAN11* and *AtIAN12*, which share high sequence similarity with each other. Moreover, the generation of overexpresser stable lines was largely accomplished with the genes (*AtIAN12* and *AtIAN8*) to be expressed from both a constitutive (35S) and an inducible (estradiol) promoter. Overexpresser line of *EYFP-AtIAN12* was also produced, and to be expressed by 35S promoter. *AtIAN11* and *EYFP-AtIAN11* were cloned and will be transformed to *Arabidopsis* as well. The overexpresser lines shall primarily be used to investigate the genes overexpression effect on the plant response to pathogen infection.

Pathogen infection assay using *in planta* bacterial (*Pst* DC3000) proliferation assay was applied on *ian11* mutant plants. Preliminary results indicate that the virulent bacteria proliferated similarly in both *ian11* mutant and wt plants during the virulent strain proliferation assay (Figure 3.25, C). In contrast, increase in plant susceptibility, monitored by the increase in bacterial proliferation, was noticed during avirulent (avrRpt2) strain proliferation assay in *ian11* mutant plants (Figure 3.25, D1). The positive controls *ndr1-1* and *pen-1* mutants showed as well increase in plant susceptibility to infection. It is considered important to mention that the disease assays on the mutants are promising but remain preliminary similarly to *nhl*s.

4.5.3 Expression analysis of *AtIAN* genes

Primarily, *AtIAN8* was discovered in plants after its isolation in *Arabidopsis thaliana* after treatment by *Ps* pv. maculicola carrying the effector *avrRpt2* (Reuber and Ausubel, 1996). In the present study, preliminary mRNA transcript quantification using real time PCR identified the induction of *AtIAN* genes by virulent *Pst* DC3000 and SA (data not shown, Master thesis: (Mwaanga, 2011)) while specifically *AtIAN8* was induced by avirulent *Pst* DC3000 (avrRpt2). Interestingly, all three *AtIAN* genes were induced by *Pst* DC3000, but not induced by flg22. These results indicate that *AtIAN*s are only induced by the signal transduction cascades leading to ETI (see 1.2.1.2).

5. Conclusions and future perspectives

In the present study the combination of both experimental and prediction methodologies allowed the identification of several novel PTS1 tripeptides and peroxisomal *Arabidopsis* proteins. Several peroxisome-targeted proteins are implicated in plant defense mechanisms based on protein annotations, domain conservation, sequence homology and microarray-based expression data. The development of the first high-accuracy prediction method for plant PTS1 proteins will be instrumental in identifying low-abundance and stress-inducible peroxisomal proteins as indicated in the present study.

Proteome and prediction methods identified 17 defense-related proteins, six of which were demonstrated to be targeted to peroxisomes (AtIAN12, NHL4, AtMIF1, AtMIF1, AtCAD7, and AtLIMDP), four are strongly indicated to be in peroxisomes (AtDRP, AtCDR1, NHL6, NHL25,), and six were found to be organelle-targeted (NHL39, NHL39H1, NHL13H1, NHLx, AtIAN8 and AtIAN11). The high number of newly identified peroxisomal proteins with predicted functions in plant defense against pathogens will be used intensively to understand the emerging evidences of peroxisome functions in defense responses and plant innate immunity. Moreover, the preliminary molecular studies on two immune-related families (NHL and AtIAN) and the peroxisomal identification of several members from both families will be instrumental in understanding peroxisome functions in innate immunity. Future goals emerging from this study are numerous and might advance our understanding of the new roles of peroxisome in plant innate immunity. In the near future, the NHL and AtIAN protein families will be further dissected to understand their roles in plant innate immunity.

Several detoxification-related proteins were investigated by *in vivo* subcellular localization studies. These studies could verify the peroxisomal identification of GR1 and its targeting signal TNL> as a novel functional PTS1. The ASC-GSH cycle functional studies were initiated in order to understand its mode of action in peroxisomes, involving heterologous expression of GR1 and DHAR1, isolation of peroxisomes for biochemical studies of the cycle in *Arabidopsis* leaf peroxisomes, and finally isolation of T-DNA homozygous mutants.

6. References

Agrios, G.N. (2005). Plant pathology. (Elsevier Academic Press. doi:SB731.A35).

Amundsen, I. (2009). In vivo subcellular targeting analysis indicates dual targeting of proteins to both mitochondria and peroxisomes. In CORE (Stavanger: Stavanger University).

Anzai, Y., Kim, H., Park, J., Wakabayashi, H., and Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. Int J Syst Evol Microbiol **50,** 1563-1589.

Arabidopsis genome initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408,** 796-815.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. Nature **415,** 977-983.

Assaad, F.F., Qiu, J.L., Youngs, H., Ehrhardt, D., Zimmerli, L., Kalde, M., Wanner, G., Peck, S.C., Edwards, H., Ramonell, K., Somerville, C.R., and Thordal-Christensen, H. (2004). The PEN1 syntaxin defines a novel cellular compartment upon fungal attack and is required for the timely assembly of papillae. Mol Biol Cell **15,** 5118-5129.

Babujee, L., Wurtz, V., Ma, C., Lueder, F., Soni, P., van Dorsselaer, A., and Reumann, S. (2010). The proteome map of spinach leaf peroxisomes indicates partial compartmentalization of phylloquinone (vitamin K1) biosynthesis in plant peroxisomes. J Exp Bot **61,** 1441-1453.

Baker, A., and Sparkes, I.A. (2005). Peroxisome protein import: some answers, more questions. Curr Opin Plant Biol **8,** 640-647.

Bari, R., and Jones, J. (2009). Role of plant hormones in plant defence responses. Plant Molecular Biology **69,** 473-488.

Bednarek, P., and Schulze-Lefert, P. (2009). Role of Plant Secondary Metabolites at the Host-Pathogen Interface. (Wiley-Blackwell).

Bednarek, P., Kwon, C., and Schulze-Lefert, P. (2010). Not a peripheral issue: secretion in plant-microbe interactions. Current Opinion in Plant Biology **13,** 378-387.

Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubsky, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R., Sanchez-Vallet, A., Molina, A., and Schulze-Lefert, P. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. Science **323,** 101-106.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340,** 783-795.

Bueso, E., Alejandro, S., Carbonell, P., Perez-Amador, M.A., Fayos, J., Belles, J.M., Rodriguez, P.L., and Serrano, R. (2007). The lithium tolerance of the Arabidopsis cat2 mutant reveals a cross-talk between oxidative stress and ethylene. Plant J **52,** 1052-1065.

Cai, Y.D., Liu, X.J., Xu, X.B., and Chou, K.C. (2002). Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. J Cell Biochem **84,** 343-348.

Cambot, M., Aresta, S., Kahn-Perles, B., de Gunzburg, J., and Romeo, P.H. (2002). Human immune associated nucleotide 1: a member of a new guanosine triphosphatase family expressed in resting T and B cells. Blood **99,** 3293-3301.

Cao, H., Li, X., and Dong, X. (1998). Generation of broad-spectrum disease resistance by overexpression of an essential regulatory gene in systemic acquired resistance. Proc Natl Acad Sci U S A **95,** 6531-6536.

Cao, H., Bowling, S.A., Gordon, A.S., and Dong, X. (1994). Characterization of an Arabidopsis Mutant That Is Nonresponsive to Inducers of Systemic Acquired Resistance. The Plant cell **6,** 1583-1592.

Century, K.S., Holub, E.B., and Staskawicz, B.J. (1995). NDR1, a locus of Arabidopsis thaliana that is required for disease resistance to both a bacterial and a fungal pathogen. Proc Natl Acad Sci U S A **92,** 6597-6601.

Century, K.S., Shapiro, A.D., Repetti, P.P., Dahlbeck, D., Holub, E., and Staskawicz, B.J. (1997). NDR1, a pathogen-induced component required for Arabidopsis disease resistance. Science **278,** 1963-1965.

Champigny, M.J., and Cameron, R.K. (2009). Action at a Distance: Long-Distance Signals in Induced Resistance. Advances in Botanical research **51,** 122-171, doi: 110.1016/S0065.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). ONE-STEP PREPARATION OF COMPETENT ESCHERICHIA-COLI - TRANSFORMATION AND STORAGE OF BACTERIAL-CELLS IN THE SAME SOLUTION. Proceedings of the National Academy of Sciences of the United States of America **86,** 2172-2175.

Ciccarelli, F.D., and Bork, P. (2005). The WHy domain mediates the response to desiccation in plants and bacteria. Bioinformatics **21,** 1304-1307.

Clay, N.K., Adio, A.M., Denoux, C., Jander, G., and Ausubel, F.M. (2009). Glucosinolate Metabolites Required for an Arabidopsis Innate Immune Response. Science **323,** 95-101.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16,** 735-743.

Coppinger, P., Repetti, P.P., Day, B., Dahlbeck, D., Mehlert, A., and Staskawicz, B.J. (2004). Overexpression of the plasma membrane-localized NDR1 protein results in enhanced bacterial disease resistance in Arabidopsis thaliana. Plant Journal **40,** 225-237.

Corpas, F.J., Barroso, J.B., and del Rio, L.A. (2001). Peroxisomes as a source of reactive oxygen species and nitric oxide signal molecules in plant cells. Trends Plant Sci **6,** 145-150.

Creelman, R.A., and Mullet, J.E. (1997). Biosynthesis and Action of Jasmonates in Plants. Annu Rev Plant Physiol Plant Mol Biol **48,** 355-381.

Crowell, D.N. (2000). Functional implications of protein isoprenylation in plants. Prog Lipid Res **39,** 393-408.

Crowell, D.N., and Huizinga, D.H. (2009). Protein isoprenylation: the fat of the matter. Trends Plant Sci **14,** 163-170.

Cserzo, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. (2002). On filtering false positive transmembrane protein predictions. Protein Eng **15,** 745-752.

Dangl, J.L., and Jones, J.D. (2001). Plant pathogens and integrated defence responses to infection. Nature **411,** 826-833.

Day, B., Dahlbeck, D., and Staskawicz, B.J. (2006). NDR1 interaction with RIN4 mediates the differential activation of multiple disease resistance pathways in Arabidopsis. The Plant cell **18,** 2782-2791.

De Duve, C., and Baudhuin, P. (1966). Peroxisomes (microbodies and related particles). Physiol Rev **46,** 323-357.

Debener, T., Lehnackers, H., Arnold, M., and Dangl, J.L. (1991). Identification and molecular mapping of a single Arabidopsis thaliana locus determining resistance to a phytopathogenic Pseudomonas syringae isolate. Plant J **1,** 289-302.

Delaney, T.P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., Gaffney, T., Gut-Rella, M., Kessmann, H., Ward, E., and Ryals, J. (1994). A central role of salicylic Acid in plant disease resistance. Science **266,** 1247-1250.

Dixon, D.P., and Edwards, R. (2009). Selective binding of glutathione conjugates of fatty acid derivatives by plant glutathione transferases. J Biol Chem **284,** 21249-21256.

Dong, X. (1998). SA, JA, ethylene, and disease resistance in plants. Curr Opin Plant Biol **1,** 316-323.

Dormann, P., Gopalan, S., He, S.Y., and Benning, C. (2000). A gene family in Arabidopsis thaliana with sequence similarity to NDR1 and HIN1. Plant Physiology and Biochemistry **38,** 789-796.

Edreva, A. (2005). Pathogenesis-related proteins: research progress in the last 15 years. General and Applied Plant Physiology **31,** 105-124.

Edwards, R., and Dixon, D.P. (2005). Plant glutathione transferases. In Gluthione Transferases and Gamma-Glutamyl Transpeptidases, pp. 169-186.

Emanuelsson, O., Elofsson, A., von Heijne, G., and Cristobal, S. (2003). In silico prediction of the peroxisomal proteome in fungi, plants and animals. J Mol Biol **330,** 443-456.

Escher, C.L., and Widmer, F. (1997). Lipid mobilization and gluconeogenesis in plants: do glyoxylate cycle enzyme activities constitute a real cycle? A hypothesis. Biol Chem **378,** 803-813.

Eubel, H., Meyer, E.H., Taylor, N.L., Bussell, J.D., O'Toole, N., Heazlewood, J.L., Castleden, I., Small, I.D., Smith, S.M., and Millar, A.H. (2008). Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of Arabidopsis cell culture peroxisomes. Plant Physiol **148,** 1809-1829.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K., and Bairoch, A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res **30,** 235-238.

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant Journal **18,** 265-276.

Flynn, C.R., Mullen, R.T., and Trelease, R.N. (1998). Mutational analyses of a type 2 peroxisomal targeting signal that is capable of directing oligomeric protein import into tobacco BY-2 glyoxysomes. Plant J **16,** 709-720.

Frank, J., Kaulfurst-Soboll, H., Rips, S., Koiwa, H., and von Schaewen, A. (2008). Comparative Analyses of Arabidopsis complex glycan1 Mutants and Genetic Interaction with staurosporin and temperature sensitive3a. Plant Physiology **148,** 1354-1367.

Fukao, Y., Hayashi, M., and Nishimura, M. (2002). Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of Arabidopsis thaliana. Plant Cell Physiol **43,** 689-696.

Fukao, Y., Hayashi, M., Hara-Nishimura, I., and Nishimura, M. (2003). Novel glyoxysomal protein kinase, GPK1, identified by proteomic analysis of glyoxysomes in etiolated cotyledons of Arabidopsis thaliana. Plant Cell Physiol **44,** 1002-1012.

Fulda, M., Shockey, J., Werber, M., Wolter, F.P., and Heinz, E. (2002). Two long-chain acyl-CoA synthetases from Arabidopsis thaliana involved in peroxisomal fatty acid beta-oxidation. Plant J **32,** 93-103.

Gabaldon, T. (2010). Peroxisome diversity and evolution. Philos Trans R Soc Lond B Biol Sci **365,** 765-773.

Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., Ward, E., Kessmann, H., and Ryals, J. (1993). Requirement of Salicylic Acid for the Induction of Systemic Acquired Resistance. Science **261,** 754-756.

Galichet, A., and Gruissem, W. (2003). Protein farnesylation in plants--conserved mechanisms but different targets. Curr Opin Plant Biol **6,** 530-535.

Galletti, R., Denoux, C., Gambetta, S., Dewdney, J., Ausubel, F.M., De Lorenzo, G., and Ferrari, S. (2008). The AtrbohD-Mediated Oxidative Burst Elicited by Oligogalacturonides in Arabidopsis Is Dispensable for the Activation of Defense Responses Effective against Botrytis cinerea. Plant Physiology **148,** 1695-1706.

Geraghty, M.T., Bassett, D., Morrell, J.C., Gatto, G.J., Jr., Bai, J., Geisbrecht, B.V., Hieter, P., and Gould, S.J. (1999). Detecting patterns of protein distribution and gene expression in silico. Proc Natl Acad Sci U S A **96,** 2937-2942.

Gerhardt, B. (1992). Fatty acid degradation in plants. Prog Lipid Res **31,** 417-446.

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu Rev Phytopathol **43,** 205-227.

Glazebrook, J., Rogers, E.E., and Ausubel, F.M. (1997). Use of Arabidopsis for genetic dissection of plant defense responses. Annu Rev Genet **31,** 547-569.

Golubkov, P.A., Johnson, W.H., Czerwinski, R.M., Person, M.D., Wang, S.C., Whitman, C.P., and Hackert, M.L. (2006). Inactivation of the phenylpyruvate tautomerase activity of macrophage migration inhibitory factor by 2-oxo-4-phenyl-3-butynoate. Bioorganic Chemistry **34,** 183-199.

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell **5,** 1003-1011.
Gomez-Gomez, L., Felix, G., and Boller, T. (1999). A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. Plant Journal **18,** 277-284.
Gopalan, S., Wei, W., and He, S.Y. (1996). hrp gene-dependent induction of hin1: a plant gene activated rapidly by both harpins and the avrPto gene-mediated signal. Plant J **10,** 591-600.
Goritschnig, S., Weihmann, T., Zhang, Y., Fobert, P., McCourt, P., and Li, X. (2008). A novel role for protein farnesylation in plant innate immunity. Plant Physiol **148,** 348-357.
Gould, S.G., Keller, G.A., and Subramani, S. (1987). Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. J Cell Biol **105,** 2923-2931.
Gould, S.J., Keller, G.A., Hosken, N., Wilkinson, J., and Subramani, S. (1989). A conserved tripeptide sorts proteins to peroxisomes. The Journal of cell biology **108,** 1657-1664.
Graham, I.A. (2008). Seed storage oil mobilization. Annu Rev Plant Biol **59,** 115-142.
Greenberg, J.T. (1997). Programmed Cell Death in Plant-Pathogen Interactions. Annu Rev Plant Physiol Plant Mol Biol **48,** 525-545.
Greenberg, J.T., and Yao, N. (2004). The role and regulation of programmed cell death in plant-pathogen interactions. Cell Microbiol **6,** 201-211.
Grubb, C.D., and Abel, S. (2006). Glucosinolate metabolism and its control. Trends Plant Sci **11,** 89-100.
Hardham, A.R., Takemoto, D., and White, R.G. (2008). Rapid and dynamic subcellular reorganization following mechanical stimulation of Arabidopsis epidermal cells mimics responses to fungal and oomycete attack. BMC Plant Biol **8,** 63.
Hawkins, J., Mahony, D., Maetschke, S., Wakabayashi, M., Teasdale, R.D., and Boden, M. (2007). Identifying novel peroxisomal proteins. Proteins **69,** 606-616.
Hayashi, M., and Nishimura, M. (2006). Arabidopsis thaliana--a model organism to study plant peroxisomes. Biochim Biophys Acta **1763,** 1382-1391.
Hayashi, M., Aoki, M., Kondo, M., and Nishimura, M. (1997). Changes in targeting efficiencies of proteins to plant microbodies caused by amino acid substitutions in the carboxy-terminal tripeptide. Plant Cell Physiol **38,** 759-768.
Hayashi, M., Toriyama, K., Kondo, M., and Nishimura, M. (1998). 2,4-Dichlorophenoxybutyric acid-resistant mutants of Arabidopsis have defects in glyoxysomal fatty acid beta-oxidation. The Plant cell **10,** 183-195.
Hayashi, M., Aoki, M., Kato, A., Kondo, M., and Nishimura, M. (1996). Transport of chimeric proteins that contain a carboxy-terminal targeting signal into plant microbodies. Plant J **10,** 225-234.
Heath, M.C. (2009). A Personal Perspective of the Last 40 Years of Plant Pathology: Emerging Themes, Paradigm Shifts and Future Promise. Annual Plant Reviews **34,** 1-15, doi: 10.1111.
Heath, M.l.C. (2000). Hypersensitive response-related death. Plant Molecular Biology **44,** 321-334.
Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M., He, K., Li, J., Schroeder, J.I., Peck, S.C., and Rathjen, J.P. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. Proc Natl Acad Sci U S A **104,** 12217-12222.
Hoagland, D., and Arnon, D. (1950). The water culture method for growing plants without soil. California Agric. Exp. Sta. Circ. **347,** 1-32.
Hoepfner, D., Schildknegt, D., Braakman, I., Philippsen, P., and Tabak, H.F. (2005). Contribution of the endoplasmic reticulum to peroxisome formation. Cell **122,** 85-95.
Holsters, M., Waele, D., Depicker, A., Messens, E., Montagu, M., and Schell, J. (1978). Transfection and transformation of Agrobacterium tumefaciens. Molecular and General Genetics MGG **163,** 181-187.
Hurt, E.C., Pesold-Hurt, B., Suda, K., Oppliger, W., and Schatz, G. (1985). The first twelve amino acids (less than half of the pre-sequence) of an imported mitochondrial protein can direct

mouse cytosolic dihydrofolate reductase into the yeast mitochondrial matrix. EMBO **4,** 2061-2068.

Innes, R.W., Bent, A.F., Kunkel, B.N., Bisgrove, S.R., and Staskawicz, B.J. (1993). Molecular analysis of avirulence gene avrRpt2 and identification of a putative regulatory sequence common to all known Pseudomonas syringae avirulence genes. J. Bacteriol. **175,** 4859-4869.

Jimenez, A., Hernandez, J.A., Del Rio, L.A., and Sevilla, F. (1997). Evidence for the Presence of the Ascorbate-Glutathione Cycle in Mitochondria and Peroxisomes of Pea Leaves. Plant Physiol **114,** 275-284.

Jones, J.D., and Dangl, J.L. (2006). The plant immune system. Nature **444,** 323-329.

Katagiri, F., Thilmony, R., and Y, H.S. (2002). The Arabidopsis thaliana-Pseudomonas syringae Interaction. The Arabidopsis Book **10.1199-tab.0039**.

Kato, A., Hayashi, M., Kondo, M., and Nishimura, M. (1996). Targeting and processing of a chimeric protein with the N-terminal presequence of the precursor to glyoxysomal citrate synthase. The Plant cell **8,** 1601-1611.

Kato, A., Takeda-Yoshikawa, Y., Hayashi, M., Kondo, M., Hara-Nishimura, I., and Nishimura, M. (1998). Glyoxysomal malate dehydrogenase in pumpkin: cloning of a cDNA and functional analysis of its presequence. Plant Cell Physiol **39,** 186-195.

Kaur, N., Reumann, S., and Hu, J. (2009). Peroxisome Biogenesis and Function. The Arabidopsis Book **7,** e0123, doi/0110.1199-tab.0123.

Keller, N.P., Turner, G., and Bennett, J.W. (2005). Fungal secondary metabolism - from biochemistry to genomics. Nat Rev Microbiol **3,** 937-947.

Khan, B.R., and Zolman, B.K. (2010). pex5 Mutants that differentially disrupt PTS1 and PTS2 peroxisomal matrix protein import in Arabidopsis. Plant Physiol **154,** 1602-1615.

Kiedrowski, S., Kawalleck, P., Hahlbrock, K., Somssich, I.E., and Dangl, J.L. (1992). Rapid activation of a novel plant defense gene is strictly dependent on the Arabidopsis RPM1 disease resistance locus. EMBO J **11,** 4677-4684.

Kim, D.-J., and Smith, S.M. (1994). Expression of a single gene encoding microbody NAD-malate dehydrogenase during glyoxysome and peroxisome development in cucumber. Plant Molecular Biology **26,** 1833-1841.

Kisaki, T., and Tolbert, N.E. (1969). Glycolate and glyoxylate metabolism by isolated peroxisomes or chloroplasts. Plant Physiol **44,** 242-250.

Koh, S., Andre, A., Edwards, H., Ehrhardt, D., and Somerville, S. (2005). Arabidopsis thaliana subcellular responses to compatible Erysiphe cichoracearum infections. Plant J **44,** 516-529.

Kragler, F., Lametschwandtner, G., Christmann, J., Hartig, A., and Harada, J.J. (1998). Identification and analysis of the plant peroxisomal targeting signal 1 receptor NtPEX5. Proc Natl Acad Sci U S A **95,** 13336-13341.

Kragt, A., Voorn-Brouwer, T., van den Berg, M., and Distel, B. (2005). Endoplasmic reticulum-directed Pex3p routes to peroxisomes and restores peroxisome formation in a Saccharomyces cerevisiae pex3Delta strain. J Biol Chem **280,** 34350-34357.

Krucken, J., Schroetel, R.M.U., Muller, I.U., Saidani, N., Marinovski, P., Benten, W.P.M., Stamm, O., and Wunderlich, F. (2004). Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. Gene **341,** 291-304.

Kuzniak, E., and Sklodowska, M. (2004). Compartment-specific role of the ascorbate-glutathione cycle in the response of tomato leaf cells to Botrytis cinerea infection. J Exp Bot **56,** 921-933.

Kuzniak, E., and Sklodowska, M. (2005). Fungal pathogen-induced changes in the antioxidant systems of leaf peroxisomes from infected tomato plants. Planta **222,** 192-200.

Kwon, C., Bednarek, P., and Schulze-Lefert, P. (2008). Secretory pathways in plant immune responses. Plant Physiology **147,** 1575-1583.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J Mol Biol **157,** 105-132.

Lee, S.B., Ham, B.K., Park, J.M., Kim, Y.J., and Paek, K.H. (2006). BnNHL18A shows a localization change by stress-inducing chemical treatments. Biochemical and Biophysical Research Communications **339,** 399-406.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. **317,** 41-72.

Leterrier, M., Corpas, F.J., Barroso, J.B., Sandalio, L.M., and del Rio, L.A. (2005). Peroxisomal monodehydroascorbate reductase. Genomic clone characterization and functional analysis under environmental stress conditions. Plant Physiol **138,** 2111-2123.

Li, J.F., Park, E., von Arnim, A.G., and Nebenfuhr, A. (2009). The FAST technique: a simplified Agrobacterium-based transformation method for transient gene expression analysis in seedlings of Arabidopsis and other plant species. Plant Methods **5,** 6.

Lingard, M.J., Monroe-Augustus, M., and Bartel, B. (2009). Peroxisome-associated matrix protein degradation in Arabidopsis. Proc Natl Acad Sci U S A **106,** 4561-4566.

Lingard, M.J., Gidda, S.K., Bingham, S., Rothstein, S.J., Mullen, R.T., and Trelease, R.N. (2008). Arabidopsis PEROXIN11c-e, FISSION1b, and DYNAMIN-RELATED PROTEIN3A cooperate in cell cycle-associated replication of peroxisomes. The Plant cell **20,** 1567-1585.

Lingner, T., Kataya, A.R., Antonicelli, G.E., Benichou, A., Nilssen, K., Chen, X.Y., Siemsen, T., Morgenstern, B., Meinicke, P., and Reumann, S. (2011). Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. The Plant cell **23,** 1556-1572.

Lipka, V., Dittgen, J., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., Landtag, J., Brandt, W., Rosahl, S., Scheel, D., Llorente, F., Molina, A., Parker, J., Somerville, S., and Schulze-Lefert, P. (2005). Pre- and postinvasion defenses both contribute to nonhost resistance in Arabidopsis. Science **310,** 1180-1183.

Lisenbee, C.S., Lingard, M.J., and Trelease, R.N. (2005). Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. Plant J **43,** 900-914.

Liu, C., Wang, T., Zhang, W.S., and Li, X. (2008). Computational identification and analysis of immune-associated nucleotide gene family in Arabidopsis thaliana. J. Plant Physiol. **165,** 777-787.

Lopez-Huertas, E., Charlton, W.L., Johnson, B., Graham, I.A., and Baker, A. (2000). Stress induces peroxisome biogenesis genes. EMBO J **19,** 6770-6777.

Luna, E., Pastor, V., Robert, J.r.m., Flors, V., Mauch-Mani, B., and Ton, J. (2011). Callose Deposition: A Multifaceted Plant Defense Response. Molecular Plant-Microbe Interactions **24,** 183-193.

Ma, C., and Reumann, S. (2008). Improved prediction of peroxisomal PTS1 proteins from genome sequences based on experimental subcellular targeting analyses as exemplified for protein kinases from Arabidopsis. J Exp Bot **59,** 3767-3779.

Ma, C., Haslbeck, M., Babujee, L., Jahn, O., and Reumann, S. (2006). Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. Plant Physiol **141,** 47-60.

Mackey, D., Holt, B.F., 3rd, Wiig, A., and Dangl, J.L. (2002). RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. Cell **108,** 743-754.

Mackey, D., Belkhadir, Y., Alonso, J.M., Ecker, J.R., and Dangl, J.L. (2003). Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell **112,** 379-389.

Marelli, M., Smith, J.J., Jung, S., Yi, E., Nesvizhskii, A.I., Christmas, R.H., Saleem, R.A., Tam, Y.Y., Fagarasanu, A., Goodlett, D.R., Aebersold, R., Rachubinski, R.A., and Aitchison, J.D. (2004). Quantitative mass spectrometry reveals a role for the GTPase Rho1p in actin organization on the peroxisome membrane. J Cell Biol **167,** 1099-1112.

Matre, P., Meyer, C., and Lillo, C. (2009). Diversity in subcellular targeting of the PP2A B'eta subfamily members. Planta **230,** 935-945.

Maurer-Stroh, S., and Eisenhaber, F. (2005). Refinement and prediction of protein prenylation motifs. Genome Biol **6,** R55.

Metraux, J.-P., Jackson, R.W., Schnettler, E., and Goldbach, R.W. (2009). Plant Pathogens as Suppressors of Host Defense. Advances in Botanical research **51,** 40-89, doi: 10.1016/S0065.

Meyer, T., Hölscher, C., Schwöppe, C., and von Schaewen, A. (2011). Alternative targeting of Arabidopsis plastidic glucose-6-phosphate dehydrogenase G6PD1 involves cysteine-dependent interaction with G6PD4 in the cytosol. The Plant Journal**,** no-no.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. The Plant cell **15,** 809-834.

Michels, P.A. (1988). Compartmentation of glycolysis in trypanosomes: a potential target for new trypanocidal drugs. Biol Cell **64,** 157-164.

Mitsuya, S., El-Shami, M., Sparkes, I.A., Charlton, W.L., Lousa, C.D., Johnson, B., and Baker, A. Salt Stress Causes Peroxisome Proliferation, but Inducing Peroxisome Proliferation Does Not Improve NaCl Tolerance in Arabidopsis thaliana. Plos One **5**.

Moschou, P.N., Sanmartin, M., Andriopoulou, A.H., Rojo, E., Sanchez-Serrano, J.J., and Roubelakis-Angelakis, K.A. (2008). Bridging The Gap Between Plant And Mammalian Polyamine Catabolism: A Novel Peroxisomal Polyamine Oxidase Responsible For A Full Back-Conversion Pathway In Arabidopsis thaliana. Plant Physiol.

Mudgett, M.B., and Staskawicz, B.J. (1999). Characterization of the Pseudomonas syringae pv. tomato AvrRpt2 protein: demonstration of secretion and processing during bacterial pathogenesis. Molecular Microbiology **32,** 927-941.

Mullen, R.T., and Trelease, R.N. (2006). The ER-peroxisome connection in plants: development of the "ER semi-autonomous peroxisome maturation and replication" model for plant peroxisome biogenesis. Biochim Biophys Acta **1763,** 1655-1668.

Mullen, R.T., Flynn, C.R., and Trelease, R.N. (2001). How are peroxisomes formed? The role of the endoplasmic reticulum and peroxins. Trends in Plant Science **6,** 256-261.

Mullen, R.T., Lee, M.S., Flynn, C.R., and Trelease, R.N. (1997). Diverse amino acid residues function within the type 1 peroxisomal targeting signal. Implications for the role of accessory residues upstream of the type 1 peroxisomal targeting signal. Plant Physiol **115,** 881-889.

Mwaanga, C. (2011). Identification and expression analysis of peroxisome-targeted defence proteins mediating innate immunity in the model plant *Arabidopsis thaliana*. In CORE (Stavanger: Stavanger University), pp. 97.

Nakai, K., and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14,** 897-911.

Narendra, S., Venkataramani, S., Shen, G., Wang, J., Pasapula, V., Lin, Y., Kornyeyev, D., Holaday, A.S., and Zhang, H. (2006). The Arabidopsis ascorbate peroxidase 3 is a peroxisomal membrane-bound antioxidant enzyme and is dispensable for Arabidopsis growth and development. Journal of Experimental Botany **57,** 3033-3042.

Nelson, B.K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. The Plant Journal **51,** 1126-1136.

Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003). Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol **328,** 581-592.

Nilssen, K. (2009). Identification of novel *Arabidopsis* proteins of peroxisomes by in vivo subcellular targeting analysis for the improvement of computational protein prediction algorithms. In CORE (Stavanger: Stavanger University), pp. 59.

Nishimura, M.T., Stein, M., Hou, B.H., Vogel, J.P., Edwards, H., and Somerville, S.C. (2003). Loss of a callose synthase results in salicylic acid-dependent disease resistance. Science **301,** 969-972.

Nito, K., Hayashi, M., and Nishimura, M. (2002). Direct Interaction and Determination of Binding Domains among Peroxisomal Import Factors in Arabidopsis thaliana. Plant and Cell Physiology **43,** 355-366.

Nyathi, Y., and Baker, A. (2006). Plant peroxisomes as a source of signalling molecules. Biochim Biophys Acta **1763,** 1478-1495.

Nürnberger, T., and Kemmerling, B. (2009). Pathogen-Associated Molecular Patterns (PAMP) and PAMP-Triggered Immunity. Annual Plant Reviews **34,** 16-47, doi: 10.1111.

Olsen, L.J. (1998). The surprising complexity of peroxisome biogenesis. Plant Mol Biol **38,** 163-189.

Orth, T., Reumann, S., Zhang, X.C., Fan, J.L., Wenzel, D., Quan, S., and Hu, J.P. (2007). The PEROXIN11 protein family controls peroxisome proliferation in Arabidopsis. The Plant cell **19,** 333-350.

Parker, J.E., Holub, E.B., Frost, L.N., Falk, A., Gunn, N.D., and Daniels, M.J. (1996). Characterization of eds1, a mutation in Arabidopsis suppressing resistance to Peronospora parasitica specified by several different RPP genes. The Plant cell **8,** 2033-2046.

Poirier, Y., Antonenkov, V.D., Glumoff, T., and Hiltunen, J.K. (2006). Peroxisomal beta-oxidation--a metabolic pathway with multiple functions. Biochim Biophys Acta **1763,** 1413-1426.

Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N.N., Kitagawa, M., Mori, H., Savchenko, A., and Yakunin, A.F. (2004). General enzymatic screens identify three new nucleotidases in Escherichia coli. Biochemical characterization of SurE, YfbR, and YjjG. J Biol Chem **279,** 54687-54694.

Purdue, P.E., and Lazarow, P.B. (2001). Peroxisome biogenesis. Annu Rev Cell Dev Biol **17,** 701-752.

Quan, S., Switzenberg, R., Reumann, S., and Hu, J. (2010). In vivo subcellular targeting analysis validates a novel peroxisome targeting signal type 2 and the peroxisomal localization of two proteins with putative functions in defense in Arabidopsis. Plant Signal Behav **5**.

Reuber, T.L., and Ausubel, F.M. (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.

Reumann, S. (2004). Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol **135,** 783-800.

Reumann, S. (2011). Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics. Proteomics **11,** 1764-1779.

Reumann, S., and Weber, A.P. (2006). Plant peroxisomes respire in the light: some gaps of the photorespiratory C2 cycle have become filled--others remain. Biochim Biophys Acta **1763,** 1496-1510.

Reumann, S., and Corpas, F.J. (2010). The Peroxisomal Ascorbate–Glutathione Pathway: Molecular Identification and Insights into Its Essential Role Under Environmental Stress Conditions. Ascorbate-glutathione pathway and stress tolerance in plants. 378, **,** doi/10.1007.

Reumann, S., Babujee, L., Ma, C., Wienkoop, S., Siemsen, T., Antonicelli, G.E., Rasche, N., Luder, F., Weckwerth, W., and Jahn, O. (2007). Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. The Plant cell **19,** 3170-3193.

Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C.G., Weber, A.P., Olsen, L.J., and Hu, J. (2009). In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol.

Rucktaschel, R., Thoms, S., Sidorovitch, V., Halbach, A., Pechlivanis, M., Volkmer, R., Alexandrov, K., Kuhlmann, J., Rottensteiner, H., and Erdmann, R. (2009). Farnesylation

of pex19p is required for its structural integrity and function in peroxisome biogenesis. J Biol Chem **284,** 20885-20896.

Rusterucci, C., Espunya, M.C., Diaz, M., Chabannes, M., and Martinez, M.C. (2007). S-nitrosoglutathione reductase affords protection against pathogens in Arabidopsis, both locally and systemically. Plant Physiol **143,** 1282-1292.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4,** 406-425.

Scheffzek, K., Ahmadian, M.R., and Wittinghofer, A. (1998). GTPase-activating proteins: helping hands to complement an active site. Trends in Biochemical Sciences **23,** 257-262.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat Genet **37,** 501-506.

Schnell, S., Demolliere, C., Van den Berk, P., and Jacobs, H. (2006). Gimap4 accelerates T-cell death. Blood **108,** 591-599.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly Specific Gene Silencing by Artificial MicroRNAs in Arabidopsis. The Plant Cell Online **18,** 1121-1133.

Schwessinger, B., and Zipfel, C. (2008). News from the frontline: recent insights into PAMP-triggered immunity in plants. Curr Opin Plant Biol **11,** 389-395.

Shah, J. (2003). The salicylic acid loop in plant defense. Current Opinion in Plant Biology **6,** 365-371.

Simöes, I., Faro, R., Bur, D., and Faro, C. (2007). Characterization of Recombinant CDR1, an Arabidopsis Aspartic Proteinase Involved in Disease Resistance. Journal of Biological Chemistry **282,** 31358-31365.

Somssich, I.E., Bollmann, J., Hahlbrock, K., Kombrink, E., and Schulz, W. (1989). Differential early activation of defense-related genes in elicitor-treated parsley cells. Plant Molecular Biology **12,** 227-234.

Stein, M., Dittgen, J., Sanchez-Rodriguez, C., Hou, B.H., Molina, A., Schulze-Lefert, P., Lipka, V., and Somerville, S. (2006). Arabidopsis PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. The Plant cell **18,** 731-746.

Subramaniam, R., Desveaux, D., Spickler, C., Michnick, S.W., and Brisson, N. (2001). Direct visualization of protein interactions in plant cells. Nat Biotechnol **19,** 769-772.

Swinkels, B.W., Gould, S.J., and Subramani, S. (1992). Targeting efficiencies of various permutations of the consensus C-terminal tripeptide peroxisomal targeting signal. FEBS Lett **305,** 133-136.

Taler, D., Galperin, M., Benjamin, I., Cohen, Y., and Kenigsbuch, D. (2004). Plant eR genes that encode photorespiratory enzymes confer resistance against disease. The Plant cell **16,** 172-184.

Tang, X., Frederick, R.D., Zhou, J., Halterman, D.A., Jia, Y., and Martin, G.B. (1996). Initiation of Plant Disease Resistance by Physical Interaction of AvrPto and Pto Kinase. Science **274,** 2060-2063.

Thaler, J., Karban, R., Ullman, D., Boege, K., and Bostock, R. (2002). Cross-talk between jasmonate and salicylate plant defense pathways: effects on several plant parasites. Oecologia **131,** 227-235.

Tolbert, N.E., Oeser, A., Yamazaki, R.K., Hageman, R.H., and Kisaki, T. (1969). A survey of plants for leaf peroxisomes. Plant Physiol **44,** 135-147.

Trezzini, G.F., Horrichs, A., and Somssich, I.E. (1993). Isolation of putative defense-related genes from Arabidopsis thaliana and expression in fungal elicitor-treated cells. Plant Mol Biol **21,** 385-389.

van den Bosch, H., Schutgens, R.B., Wanders, R.J., and Tager, J.M. (1992). Biochemistry of peroxisomes. Annu Rev Biochem **61,** 157-197.

Varet, A., Hause, B., Hause, G., Scheel, D., and Lee, J. (2003). The Arabidopsis NHL3 gene encodes a plasma membrane protein and its overexpression correlates with increased resistance to Pseudomonas syringae pv. tomato DC3000. Plant Physiol **132,** 2023-2033.

Varet, A., Parker, J., Tornero, P., Nass, N., Nurnberger, T., Dangl, J.L., Scheel, D., and Lee, J. (2002). NHL25 and NHL3, two NDR1/HIN1-like genes in Arabidopsis thaliana with potential role(s) in plant defense. Molecular Plant-Microbe Interactions **15,** 608-616.

Veenhuis, M., Van Dijken, J.P., and Harder, W. (1983). The significance of peroxisomes in the metabolism of one-carbon compounds in yeasts. Adv Microb Physiol **24,** 1-82.

Vlot, A.C., Klessig, D.F., and Park, S.W. (2008). Systemic acquired resistance: the elusive signal(s). Curr Opin Plant Biol **11,** 436-442.

Wanders, R.J., and Waterham, H.R. (2006). Biochemistry of mammalian peroxisomes revisited. Annu Rev Biochem **75,** 295-332.

Wang, Z., and Li, X. (2009). IAN/GIMAPs are conserved and novel regulators in vertebrates and angiosperm plants. Plant Signal Behav **4,** 165-167.

Weigel, D., and Brook, J.G. (2002). Arabidopsis: a laboratory manual Cold spring harbor laboratory press**,** doi: 0-87969-87572-87962.

Wierzbicki, A.S. (2007). Peroxisomal disorders affecting phytanic acid alpha-oxidation: a review. Biochem Soc Trans **35,** 881-886.

Woodward, A.W., and Bartel, B. (2005). Auxin: regulation, action, and interaction. Ann Bot **95,** 707-735.

Xia, Y., Suzuki, H., Borevitz, J., Blount, J., Guo, Z., Patel, K., Dixon, R.A., and Lamb, C. (2004). An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. EMBO J **23,** 980-988.

Yan, L., and Robert, L. (2008). Web-Based *Arabidopsis* Functional and Structural Genomics Resources. The Arabidopsis Book **10.1199-tab.0118**.

Yang, H., Shi, Y., Liu, J., Guo, L., Zhang, X., and Yang, S. (2010). A mutant CHS3 protein with TIR-NB-LRR-LIM domains modulates growth, cell death and freezing tolerance in a temperature-dependent manner in Arabidopsis. The Plant Journal **63,** 283-296.

Zhang, X., and Hu, J. (2009). Two small protein families, DYNAMIN-RELATED PROTEIN3 and FISSION1, are required for peroxisome fission in Arabidopsis. The Plant Journal **57,** 146-159.

Zhang, X.C., and Hu, J.P. (2008). FISSION1A and FISSION1B Proteins Mediate the Fission of Peroxisomes and Mitochondria in Arabidopsis. Molecular Plant **1,** 1036-1047.

Zhang, X.C., and Hu, J.P. (2010). The Arabidopsis Chloroplast Division Protein DYNAMIN-RELATED PROTEIN5B Also Mediates Peroxisome Division. The Plant cell **22,** 431-442.

Zhang, Y., Liu, Z., Wang, L., Zheng, S., Xie, J., and Bi, Y. (2010). Sucrose-induced hypocotyl elongation of Arabidopsis seedlings in darkness depends on the presence of gibberellins. J Plant Physiol **167,** 1130-1136.

Zheng, M.S., Takahashi, H., Miyazaki, A., Hamamoto, H., Shah, J., Yamaguchi, I., and Kusano, T. (2004). Up-regulation of Arabidopsis thaliana NHL10 in the hypersensitive response to Cucumber mosaic virus infection and in senescing leaves is controlled by signalling pathways that differ in salicylate involvement. Planta **218,** 740-750.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol **136,** 2621-2632.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J.D.G., Felix, G., and Boller, T. (2004). Bacterial disease resistance in Arabidopsis through flagellin perception. Nature **428,** 764-767.

Zolman, B.K., Silva, I.D., and Bartel, B. (2001). The Arabidopsis pxa1 Mutant Is Defective in an ATP-Binding Cassette Transporter-Like Protein Required for Peroxisomal Fatty Acid beta -Oxidation. Plant Physiol. **127,** 1266-1278.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K.J. (2008). Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome. PLoS ONE **3,** e1994.

7. Appendix

Table 2.4: Primers used for cloning and genotyping

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
SR491f	At3g51660_AtMIF1	G50544/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGCCTTGTCTTTACATTAC	NotI
SR492r	At3g51660_AtMIF1	G50544/ABRC	pCAT-EYFP	CAAGTCTAGAGCTAAAGTTTAGAAGGAAGAG	XbaI
AK27F	At5g01650.1 _AtMIF2	U17152/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGCCGTGCCTCAACCTCTCC	NotI
AK28R	At5g01650.1 _AtMIF2	U17152/ABRC	pCAT-EYFP	CAAGTCTAGAGTTAAAGAGTCGCCCCGTTCCA	XbaI
AK3F	At4g14930_AtSurE	U25020/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGAGATTGACGGTGGAGAT	NotI
AK4R	At4g14930_AtSurE	U25020/ABRC	pCAT-EYFP	CAAGTCTAGAGTCAAAGGGATGAGGAGGAGCA	XbaI
AK183R	EYFP-PTD (AtSurE, SSL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCAAAGGGATGAGGAGGAGCATGACTGGTTTGTCTTGTACAGCTCGTCC ATGCC	XbaI
AK5F	At5g33340 _AtCDR1	U85644/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCCTCTCTATTCTCTTCA	NotI
AK6R	At5g33340 _AtCDR1	U85644/ABRC	pCAT-EYFP	CAAGTCTAGAGCTACATCTTTGCACAATCTGT	XbaI
AK9F	At5g17890.1_AtLIMDP	pda07886/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCTATAGCATTAATAGAGTCAAAG	NotI
AK10R	At5g17890.1_AtLIMDP	pda07886/RIKEN	pCAT-EYFP	CAAGTCTAGAGTCATAACTTTGAATATTGTGG	XbaI
AK62F	AtLIMDP_SDM (ΔT)	pCAT-AtLIMDP	pCAT-EYFP	GTGTTAGGCTATATGTGGTTGGAGTGTCAGACATACGTTTTTG	
AK63R	AtLIMDP_SDM (ΔT)	pCAT-AtLIMDP	pCAT-EYFP	CAAAAACGTATGTCTGACACTCCAACCACATATAGCCTAACAC	
AK57F	AT1G58807.2_AtDRP	pda20094/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCTGGTTGGAAAGTGTTGCTTACTTCT	NotI
AK58R	AT1G58807.2_AtDRP	pda20094/RIKEN	pCAT-EYFP	CAAGCCGCGGCTAGAGTCGGCAATGGATCTGAATATCGAGAGT	SacII
AK182R	EYFP-PTD (DRP, CRL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCAGAGTCGGCAATGGATCTGAATATCGAGAGTCTTGTACAGCTCGTCCA TGCC	XbaI
AK60F	At4g37980.2_AtCAD7	pda01912/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCT ATGGGAAAGGTTCTTGAGAAG	NotI
AK61R	At4g37980.2_AtCAD7	pda01912/RIKEN	pCAT-EYFP	CAAGCCGCGGTTAAAGATGACTGACAAATAGGTTCATACCAAAGATGAGAGG	SacII
AK184R	YFP-PTD (CAD7, SHL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCAAAGATGACTGACAAATAGGTTCATACCAAACTTGTACAGCTCGTCC ATGCC	XbaI
AK1F	AT4G09940_AtIAN12 (CIIM>)	DQ056647/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGTTTTCAGAATCTCTCCCA	NotI
AK2R	AT4G09940_AtIAN12 (CIIM>)	DQ056647/ABRC	pCAT-EYFP	CAAGCCGCGGTCACATAATGATGCACCACTG	SacII
AK73R	AT4G09940_AtIAN12 (CSKL>)	DQ056647/ABRC	pCAT-EYFP	CAAGCCGCGGTCACAATTTGGAGCACCACTGCTTGGATTT	SacII
AK65R	AT4G09940_AtIAN12 (ΔIIM>)	DQ056647/ABRC	pCAT-EYFP	CAAGCCGCGGTCAGCACCACTGCTTGGATTT	SacII

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
AK164R	AT4G09940_AtIAN12 (GIIM>)	DQ056647/ABRC	pCAT-EYFP	CAAGCCGCGGTCACATAATGATGCCCCACTG	SacII
AK104F	AT4G09940_AtIAN12	DQ056647/ABRC	pBA002 & pER10	ACTTTAATTAACATGTTTTCAGAATCTCTCCCA	PacI
AK105R	AT4G09940_AtIAN12	DQ056647/ABRC	pBA002 & pER10	AGACTAGTTCACATAATGATGCACCACTG	SpeI
AK19R	EYFP-6aa-CIIM>_AT4G09940	pCAT-EYFP	pCAT	TATGTCTAGAGTCACATAATGATGCACCACTGCTTGGATTTTTGCTTGTACAGCTCGTCCATGCC	XbaI
AK74R	EYFP-6aa-CSKL>_AT4G09940	pCAT-EYFP	pCAT	TATGTCTAGAGTCACAATTTGGAGCACCACTGCTTGGATTTTTGCTTGTACAGCTCGTCCATGCC	XbaI
AK75R	EYFP-6aa-AIIM>_AT4G09940	pCAT-EYFP	pCAT	TATGTCTAGAGTCACATAATGATGGCCCACGCCTTGGATTTTTGCTTGTACAGCTCGTCCATGCC	XbaI
AK64R	EYFP-6aa-CIIW>_AT4G09940	pCAT-EYFP	pCAT	TATGTCTAGAGTCACCAAATGATGCACCACTGCTTGGATTTTTGCTTGTACAGCTCGTCCATGCC	XbaI
AK33F	At4g09930_AtIAN11	GC103086/ABRC	pCAT-EYFP	AAGACTGCGGCCGCT ATGGGTGGAGGACTCGTAGAA	NotI
AK34R	At4g09930_AtIAN11	GC103086/ABRC	pCAT-EYFP	CAAGTCTAGAG TCAAAGAATGATGCAACCTTG	XbaI
AK35R	EYFP-6aa-CIIL>_At4g09930	pCAT-EYFP	pCAT	TATGTCTAGAGTCAAAGAATGATGCAACCTTGATCCCTTTTCTCCTTGTACAGCTCGTCCATGCC	XbaI
AK108F	At4g09930_AtIAN11	GC103086/ABRC	pBA002 & pER10	ACTTTAATTAACATGGGTGGAGGACTCGTAGAA	PacI
AK109R	At4g09930_AtIAN11	GC103086/ABRC	pBA002 & pER10	AGACTAGTTCAAAGAATGATGCAACCTTG	SpeI
AK36F	At1g33960_AtIAN8	pda15002/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCT ATGGCCAACGATCAGAAGAAT	NotI
AK37R	At1g33960_AtIAN8	pda15002/RIKEN	pCAT-EYFP	CAAGTCTAGAG TCAGAGAATGCTGCACTGCTG	XbaI
AK38R	EYFP-6aa-CSIL>_At1g33960	pCAT-EYFP	pCAT	TATGTCTAGAGTCAGAGAATGCTGCACTGCTGACGGCTGAGCATCTTGTACAGCTCGTCCATGCC	XbaI
AK106F	At1g33960_AtIAN8	pda15002/RIKEN	pBA002 & pER10	AAGACTCTCGAGATGGCCAACGATCAGAAGAAT	XhoI
AK107R	At1g33960_AtIAN8	pda15002/RIKEN	pBA002 & pER10	AGACTAGTTCAGAGAATGCTGCACTGCTG	SpeI
	AtIAN8_amiRNA		pBA002 & pER10	TATAAAAACGTGTCGCCTCAC	
AK110	I miR-s	pRS300		GATATAAAAACGTGTCGCCTCACTCTCTCTTTTGTATTCC	
AK111	II miR-a	pRS300		GAGTGAGGCGACACGTTTTTATATCAAAGAGAATCAATGA	
AK112	III miR*s	pRS300		GAGTAAGGCGACACGATTTTATTTCACAGGTCGTGATATG	
AK113	IV miR*a	pRS300		GAAATAAAATCGTGTCGCCTTACTCTACATATATATTCCT	

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
	AtIAN12_amiRNA		pBA002 & pER10	TATCTTTAATGCAAAAGGCGC	
AK122	I miR-s	pRS300		GATATCTTTAATGCAAAAGGCGCTCTCTCTTTTGTATTCC	
AK123	II miR-a	pRS300		GAGCGCCTTTTGCATTAAAGATATCAAAGAGAATCAATGA	
AK124	III miR*s	pRS300		GAGCACCTTTTGCATAAAAGATTTCACAGGTCGTGATATG	
AK125	IV miR*a	pRS300		GAAATCTTTTATGCAAAAGGTGCTCTACATATATATTCCT	
	AtIAN12+AtIAN11_ amiRNA		pBA002 & pER10	TAGAATGCTATTCCGTGTCGC	
AK130	I miR-s	pRS300		GATAGAATGCTATTCCGTGTCGCTCTCTCTTTTGTATTCC	
AK131	II miR-a	pRS300		GAGCGACACGGAATAGCATTCTATCAAAGAGAATCAATGA	
AK132	III miR*s	pRS300		GAGCAACACGGAATACCATTCTTTCACAGGTCGTGATATG	
AK133	IV miR*a	pRS300		GAAAGAATGGTATTCCGTGTTGCTCTACATATATATTCCT	
AK7F	At1g54540_NHL4	PENTR221-AT1G54540/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGGAGATCAACAAAAAATT	NotI
AK8R	At1g54540_NHL4	PENTR221-AT1G54540/ABRC	pCAT-EYFP	CAAGGAGCTCTCAGAGTTTGGCCTTAAAACT	SacI
AK134F	At1g54540_NHL4	PENTR221-AT1G54540/ABRC	pBA002 & pER10	AAGACTCTCGAGATGGGAGATCAACAAAAAATT	XhoI
AK135R	At1g54540_NHL4	PENTR221-AT1G54540/ABRC	pBA002 & pER10	AGACTAGTTCAGAGTTTGGCCTTAAAACT	SpeI
AK180F	EYFP-NHL4	PENTR221-AT1G54540/ABRC	pBA002 & pER10	ACTTTAATTAACATGGGAGATCAACAAAAAATT	PacI
AK181R	EYFP-PTD (NHL4, AKL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCAGAGTTTGGCCTTAAAACTGCAATCACTAGCCTTGTACAGCTCGTCCATGCC	XbaI
AK43F	At1g65690_NHL6	Senescent leaves_mRNA	pCAT-EYFP	AAGACTGCGGCCGCTATGTCTCAACACCAAAAAATCTATCCGGTCCAAG	NotI
AK44R	At1g65690_NHL6	Senescent leaves_mRNA	pCAT-EYFP	CAAGCCGCGG CTATAACCTAAGACGAAATTTGCAACT	SacII
AK45R	EYFP-PTD (NHL6, LRL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCATAACCTAAGACGAAATTTGCAACTACTACTCTTGTACAGCTCGTCCATGCC	XbaI
AK136F	At1g65690_NHL6	To be repeated	pBA002 & pER10	ACTTTAATTAACATGTCTCAACACCAAAAAATCTAT	PacI
AK137R	At1g65690_NHL6	To be repeated	pBA002 & pER10	AGACTAGTCTATAACCTAAGACGAAATTTGCA	SpeI
AK46F	At5g36970_NHL25	SA_Sprayed_Leaves_mRNA	pCAT-EYFP	AAGACTGCGGCCGCTATGTCCGATCACCAGAAAATTCATCCGGTGAGCG	NotI

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
AK47R	At5g36970_NHL25	SA_Sprayed_Leaves_mRNA	pCAT-EYFP	CAAGTCTAGAG TTATAGTCTAAACCTGTATTTGCAGTT	XbaI
AK48R	EYFP-PTD (NHL25, FRL>)	pCAT-EYFP	pCAT	TATGTCTAGAGTCATAGTCTAAACCTGTATTTGCAGTTACTACTCTTGTACAGCTCGTCCATGCC	XbaI
AK138F	At5g36970_NHL25	SA_Sprayed_Leaves_mRNA	pBA002 & pER10	AAGACTCTCGAGATGTCCGATCACCAGAAAATTCAT	XhoI
AK139R	At5g36970_NHL25	SA_Sprayed_Leaves_mRNA	pBA002 & pER10	AGACTAGTTTATAGTCTAAACCTGTATTTGCA	SpeI
AK49F	At5g21130_NHL13H1	Genomic DNA	pCAT-EYFP	AAGACTGCGGCCGCT ATGACGGTCGAGAAACCACAA	NotI
AK50R	At5g21130_NHL13H1	Genomic DNA	pCAT-EYFP	CAAGTCTAGAG TTACAACAGGCTCAAGCCCGT	XbaI
AK51F	At3g54200_NHL39	pda19744/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCT ATGAGTGATTTTTCAATCAAA	NotI
AK52R	At3g54200_NHL39	pda19744/RIKEN	pCAT-EYFP	CAAGTCTAGAG TTATAACTTAGTCGAATACTT	XbaI
AK53F	At3g05975_NHL39H1	DQ446637/ABRC	pCAT-EYFP	AAGACTGCGGCCGCT ATGTCCAAGCGACGCATTTGC	NotI
AK54R	At3g05975_NHL39H1	DQ446637/ABRC	pCAT-EYFP	CAAGTCTAGAG TTACAGCTTAGTTTTGAGATC	XbaI
AK55F	At1g08160_NHLx	G11858/ABRC	pCAT-EYFP	AAGACTGCGGCCGCT ATGGTGCCTCCAAACCCAGCC	NotI
AK56R	At1g08160_NHLx	G11858/ABRC	pCAT-EYFP	CAAGTCTAGAG CTAAAGACGAGTTTTGCATAA	XbaI
	NHL4_amiRNA		pBA002 & pER10	TTTCGTTGGGATTACGCGCTA	
AK140	I miR-s	pRS300		GATTTCGTTGGGATTACGCGCTATCTCTCTTTTGTATTCC	
AK141	II miR-a	pRS300		GATAGCGCGTAATCCCAACGAAATCAAAGAGAATCAATGA	
AK142	III miR*s	pRS300		GATAACGCGTAATCCGAACGAATTCACAGGTCGTGATATG	
AK143	IV miR*a	pRS300		GAATTCGTTCGGATTACGCGTTATCTACATATATATTCCT	
	NHL6_amiRNA		pBA002 & pER10	TTATAGTCACGTTAAAAGCCC	
AK148	I miR-s	pRS300		GATTATAGTCACGTTAAAAGCCCTCTCTCTTTTGTATTCC	
AK149	II miR-a	pRS300		GAGGGCTTTTAACGTGACTATAATCAAAGAGAATCAATGA	
AK150	III miR*s	pRS300		GAGGACTTTTAACGTCACTATATTCACAGGTCGTGATATG	
AK151	IV miR*a	pRS300		GAATATAGTGACGTTAAAAGTCCTCTACATATATATTCCT	
	NHL25_amiRNA		pBA002 & pER10	TTATGGTAACGTTAAATCCGG	
AK156	I miR-s	pRS300		GATTATGGTAACGTTAAATCCGGTCTCTCTTTTGTATTCC	
AK157	II miR-a	pRS300		GACCGGATTTAACGTTACCATAATCAAAGAGAATCAATGA	
AK158	III miR*s	pRS300		GACCAGATTTAACGTAACCATATTCACAGGTCGTGATATG	
AK159	IV miR*a	pRS300		GAATATGGTTACGTTAAATCTGGTCTACATATATATTCCT	
SR476f	At3g24170_GR1	G25518/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCGAGGAAGATGCTT	NotI

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
SR477r	At3g24170_GR1	G25518/ABRC	pCAT-EYFP/pMAL.c2X	CAAGTCTAGAGTCATAGATTTGTCTTAGG	XbaI
SR478r	EYFP-7aa-TNL> (AtGR1)	pCAT-EYFP	pCAT	TATGTCTAGAGTCATAGATTTGTCTTAGGTTTGGGTTTGTGGGCCTTGTACAGCTCGTCCATGCC	XbaI
SR481f	At1g19570_DHAR1	pda00270/RIKEN	pCAT-EYFP	AAGACTGCGGCCGCTATGGCTCTGGAAATCTGTGT	NotI
SR482R	At1g19570_DHAR1	pda00270/RIKEN	pCAT-EYFP/pMAL.c2X	CAAGTCTAGAGTCAAGGGTTAACCTTGGG AG	XbaI
AK66F	At1g19570_DHAR1	pda00270/RIKEN	NS-EYFP	GTCACCATGGCT CTG GAA ATC TGT GT	NcoI
AK67R	At1g19570_DHAR1	pda00270/RIKEN	NS-EYFP	GAGCTCCATGGAAGGGTTAACCTTGGGAGC	NcoI
AK68R	DHAR1 (47aa)	pda00270/RIKEN	NS-EYFP	GAGCTCCATGGAGTCAGAGAGGTTAATCAGATGGAT	NcoI
AK90F	DHAR1 (47aa)_SDM (R to L)	DHAR1 (47aa)	NS-EYFP	GACTGTCCGTTCAGCCAATTGGCTCTTCTCACACTCGAG	
AK91R	DHAR1 (47aa)_SDM (R to L)	DHAR1 (47aa)	NS-EYFP	CTCGAGTGTGAGAAGAGCCAATTGGCTGAACGGACAGTC	
SR483F	At1g02920_GSTF7	U16241/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCAGGAATCAAAGTTTT	NotI
SR484R	At1g02920_GSTF7	U16241/ABRC	pCAT-EYFP	CAAGTCTAGAgTTAAAGAACCTTCTTAGCAG	XbaI
SR485F	At2g30870_GSTF10	U17031/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGTGTTGACAATCTATGC	NotI
SR486R	At2g30870_GSTF10	U17031/ABRC	pCAT-EYFP	CAAGTCTAGAGTTAAACAGGTAGTGAGTACT	XbaI
SR487F	At1g78380_GSTU19	U12572/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCGAACGAGGTGATTCT	NotI
SR488R	At1g78380_GSTU19	U12572/ABRC	pCAT-EYFP	CAAGTCTAGAGTTACTCAGGTACAAATTTCT	XbaI
SR489F	At1g78370_GSTU20	U17780/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCGAACCTACCGATTCT	NotI
SR490R	At1g78370_GSTU20	U17780/ABRC	pCAT-EYFP	CAAGTCTAGAGTCAGAGATTGTTCTTCCTAT	XbaI
SR479F	At5g43940_HMGDH	pda17160/ABRC	pCAT-EYFP	AAGACTGCGGCCGCTATGGCGACTCAAGGTCAGGT	NotI
SR480R	At5g43940_HMGDH	pda17160/ABRC	pCAT-EYFP	CAAGTCTAGAGTCATTTGCTGGTATCGAGGA	XbaI
AK18F	At5g41210_GSTT1	pGEMT-GSTT1	pQE-31	ACTGGATCCCATGATGAAGCTCAAAGTGTAT	BamHI
AK19F	At5g41210_GSTT1	pGEMT-GSTT1	pMAL-c2X	ACTGGATCCATGATGAAGCTCAAAGTGTAT	BamHI
AK20R	At5g41210_GSTT1	pGEMT-GSTT1	pQE-31/pMAL-c2X	CAAGTCGACTTAGATCTTGGATTGAAGACC	SalI
AK21F	At3g24170_GR1	G25518/ABRC	pQE-31	ACTGGATCCCATGGCGAGGAAGATGCTTGTT	BamHI
AK22F	At3g24170_GR1	G25518/ABRC	pMAL.c2X	ACTGGATCCATGGCGAGGAAGATGCTTGTT	BamHI
AK23R	At3g24170_GR1	G25518/ABRC	pQE-31	AAGGAGCTCTCATAGATTTGTCTTAGG	SacI
AK24F	At1g19570_DHAR1	pda00270/RIKEN	pQE-31	ACTGAGCTCCATGGCTCTGGAAATCTGTGTG	SacI
AK25F	At1g19570_DHAR1	pda00270/RIKEN	pMAL.c2X	ACTGAATTCATGGCTCTGGAAATCTGTGTG	EcoRI

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
AK26R	At1g19570_DHAR1	pda00270/RIKEN	pQE-31	CAAGGTCGACTCAAGGGTTAACCTTGGGAG	SalI
AK162F	EYFP	pCAT-EYFP	pGEMT-Easy	AAGACTGTCGACATGGTGAGCAAGGGCGAGGAG	SalI
AK163R	EYFP	pCAT-EYFP	pGEMT-Easy	TGCACTAGTTCCGTTAATTAACTTGTACAGCTCGTCCAT	PacI-SpeI
pRS300-A	pRS300 vector specific			CTGCAAGGCGATTAAGTTGGGTAAC	
pRS300-B	pRS300 vector specific			GCGGATAACAATTTCACACAGGAAACAG	
pER10/F	pER10 vector specific			GTGGTAATGCCATGTAATATGCTCG	
pER10/R	pER10 vector specific			ATACTCAAACTTAGTAGGATTCTGGTGTG	
pBA002/F	pBA002 vector specific			CGTCTTCAAAGCAAGTGGATTGATG	
pBA002/R	pBA002 vector specific			TGCTTAACGTAATTCAACAACAGAAATTATA	
SR194F	pCAT-N-terminal			GCATTCTACTTCTATTGCAGC	
SR320r	pCAT-C-terminal			CCTTATCTGGGAACTACTCAC	
SR321F	pCAT-downstream_EYFP			ACTACCTGAGCTACCAGTCC	
LBa1	T-DNA specific			TGGTTCACGTAGTGGGCCATCG	
LB1-SAIL	T-DNA specific			GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC	
AK94LP	NHL4 (SAIL_681_E12)_LP			TGGCCTTAAAACTGCAATCAC	
AK95RP	NHL4 (SAIL_681_E12)_RP			ACGGGTTGTTGCTGAACATAG	
AK78LP	NHL6 (SALK_148523)_LP			TGGTAAAATTTTGGCAACGAC	
AK79RP	NHL6 (SALK_148523)_RP			AATCTATCCGGTCCAAGATCC	
AK80LP	NHL25 (SALK_113216)_LP			GGCAAAAACATACGGATTGTG	
AK81RP	NHL25 (SALK_113216)_RP			GGTTACAGCTAACCCGGTTTC	
AK82LP	NHL13H1 (SALK_080000)_LP			TGCAATCACGTCCTAATCTCC	
AK83RP	NHL13H1 (SALK_080000)_RP			AAAGCCCATCAAGGCATAAAC	

Primer	Construct name	Template	Dest. vector	Nucleotide sequence (5'--3')	R. E.
AK84LP	NHL39 (SAIL_204_E02)_LP			CACACGAAATTAGGCAAAAGC	
AK85RP	NHL39 (SAIL_204_E02)_RP			CTGCGTTTCAGAGAGTAACCG	
AK86LP	NHL39H1 (SAIL_1213_B03)_LP			AACGAGTCAAACTTTAGGTGGC	
AK87RP	NHL39H1 (SAIL_1213_B03)_RP			AAGAACAGCGATCAAGAGCAC	
AK88LP	AtIAN11 (SAIL_404_H08)_LP			CCTCAAGCAATGTGGCAATAG	
AK89RP	AtIAN11 (SAIL_404_H08)_RP			GCTGCTTGTCCTTTTGACTTG	
AK100LP	LIMDP (SALK_024264)_LP			TTGAAGATTTCTTGGCAGGTG	
AK101RP	LIMDP (SALK_024264)_RP			GTTGTTTTTCCTTTCTTGGGC	
AK102LP	AtSurE (SALK_037615)_LP			CAGTTCCAGAATAGACGCTGG	
AK103RP	AtSurE (SALK_037615)_RP			TTTGGTATACGATCGAATCGC	
SR563LP	AtGR1 (SALK_105794C)_LP			TATCGATCGGGTTTGTTTTTG	
SR564RP	AtGR1 (SALK_105794C)_RP			GTTGCGGAAAAATATCAATGC	
SR565LP	DHAR1 (SALK_005382.46.25.x)_LP			ATGTCGTTTCGTATCGTCGTC	
SR566RP	DHAR1 (SALK_005382.46.25.x)_RP			TTCTCAAAAGAGTCGAGCGAG	

Paper I

The Plant Cell, Vol. 23: 1556–1572, April 2011, www.plantcell.org

Identification of Novel Plant Peroxisomal Targeting Signals by a Combination of Machine Learning Methods and in Vivo Subcellular Targeting Analyses [W]

Thomas Lingner,[a,b] Amr R. Kataya,[b] Gerardo E. Antonicelli,[b,c] Aline Benichou,[b] Kjersti Nilssen,[b] Xiong-Yan Chen,[b] Tanja Siemsen,[c] Burkhard Morgenstern,[a] Peter Meinicke,[a] and Sigrun Reumann[b,c,1]

[a] Georg-August University of Goettingen, Institute for Microbiology, Department of Bioinformatics, D-37077 Goettingen, Germany

[b] Centre for Organelle Research, University of Stavanger, N-4021 Stavanger, Norway

[c] Georg-August-University of Goettingen, Department of Plant Biochemistry, D-37077 Goettingen, Germany

In the postgenomic era, accurate prediction tools are essential for identification of the proteomes of cell organelles. Prediction methods have been developed for peroxisome-targeted proteins in animals and fungi but are missing specifically for plants. For development of a predictor for plant proteins carrying peroxisome targeting signals type 1 (PTS1), we assembled more than 2500 homologous plant sequences, mainly from EST databases. We applied a discriminative machine learning approach to derive two different prediction methods, both of which showed high prediction accuracy and recognized specific targeting-enhancing patterns in the regions upstream of the PTS1 tripeptides. Upon application of these methods to the *Arabidopsis thaliana* genome, 392 gene models were predicted to be peroxisome targeted. These predictions were extensively tested in vivo, resulting in a high experimental verification rate of *Arabidopsis* proteins previously not known to be peroxisomal. The prediction methods were able to correctly infer novel PTS1 tripeptides, which even included novel residues. Twenty-three newly predicted PTS1 tripeptides were experimentally confirmed, and a high variability of the plant PTS1 motif was discovered. These prediction methods will be instrumental in identifying low-abundance and stress-inducible peroxisomal proteins and defining the entire peroxisomal proteome of *Arabidopsis* and agronomically important crop plants.

INTRODUCTION

One of the major events that occurred during evolution was the subdivision of eukaryotic cells into membrane-enclosed subcellular compartments to optimize physiological functions. Most organellar proteins are encoded in the nucleus, translated on cytoplasmic ribosomes, and targeted to their subcellular destination by small compartment-specific targeting peptides attached to or located within the mature polypeptide (Pain et al., 1991; Schnell and Hebert, 2003). Revealing the subcellular localization of unknown proteins is of major importance for inferring protein function. To understand compartmentalization of metabolic and signal transduction networks, the proteomes of cell organelles must be defined in their full complexity. This is a challenging task using experimental approaches. The most abundant proteins of eukaryotic cell organelles have generally been identified, by classical protein chemistry or forward or reverse genetics. However, most low-abundance proteins of cell organelles have remained unidentified to date. Protein targeting prediction from genome sequences has emerged as a central tool in the postgenomic era to define organellar proteomes and to understand metabolic and regulatory networks (Schneider and Fechner, 2004; Nair and Rost, 2008; Mintz-Oron et al., 2009; Mitschke et al., 2009).

Peroxisomes are small, ubiquitous eukaryotic organelles that mediate a wide range of oxidative metabolic activities. Plant peroxisomes are essential for lipid metabolism, photorespiration, and hormone biosynthesis and metabolism, and they play pivotal roles in plant responses to abiotic and biotic stresses (Lopez-Huertas et al., 2000; Hayashi and Nishimura, 2003; Lipka et al., 2005; Nyathi and Baker, 2006; Reumann and Weber, 2006; Kaur et al., 2009). Soluble matrix proteins of peroxisomes are imported directly from the cytosol (Purdue and Lazarow, 2001). Apart from a few exceptions, proteins are targeted to the peroxisome matrix by a conserved peroxisome targeting signal of either type 1 (PTS1) or type 2 (PTS2).

Prediction methods such as PeroxiP (www.bioinfo.se/PeroxiP/) and the PTS1 predictor (mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp) and databases such as PeroxisomeDB (www.peroxisomedb.org) and AraPerox (www3.uis.no/araperoxv1) have been developed, mainly for metazoa, to predict and assemble PTS1 proteins from genomic sequences (Emanuelsson et al., 2003; Neuberger et al., 2003a, 2003b; Reumann, 2004; Reumann et al., 2004; Bodén and Hawkins, 2005; Hawkins et al., 2007; Schlüter et al., 2010). PTS1

[1] Address correspondence to sigrun.reumann@uis.no.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Sigrun Reumann (sigrun.reumann@uis.no).

[W] Online version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.111.084095

tripeptides can be roughly divided into two groups: major (canonical) and minor (noncanonical) PTS1s. Major PTS1s (e.g., SKL>, ARL>, and PRL>; ">" indicates the C-terminal end of a peptide) are the predominant signals of high-abundance proteins and are ubiquitous to most eukaryotes, providing stand-alone signals that are sufficient for peroxisome targeting. Proteins with major PTS1s can often be predicted to be peroxisomal, solely based on the PTS1 tripeptide (Reumann, 2004), or by prediction tools developed for other kingdoms and considering extended PTS1 domains (e.g., the PTS1 predictor for metazoa; Neuberger et al., 2003a, 2003b). By contrast, minor PTS1s, including the most recently discovered noncanonical PTS1s (e.g., SSI>, ASL>, and SLM> for plants; Reumann et al., 2007, 2009), are generally restricted to a few, preferentially low-abundance (weakly expressed), peroxisomal proteins and are often kingdom specific. These tripeptides alone generally represent weak signals that require auxiliary targeting-enhancing patterns (e.g., basic residues) for functionality, which are located immediately upstream of the tripeptide. Such enhancer patterns have been partially defined for metazoa (Neuberger et al., 2003a), but they appear to differ between kingdoms. Consequently, prediction tools developed for metazoa generally fail to correctly predict plant peroxisomal proteins with noncanonical PTS1 tripeptides (e.g., see Results).

The accuracy of prediction algorithms essentially relies on the size, quality, and diversity of the underlying data set of example sequences that is used for model training. Despite 40 years of peroxisome research, the number of known PTS1 proteins has remained rather low for most model organisms, and this has severely limited the size of previous training data sets to 90 to 300 sequences (Emanuelsson et al., 2003; Bodén and Hawkins, 2005; Hawkins et al., 2007). Additionally, former data sets could not reflect the natural diversity of PTS1 protein sequences and tripeptides due to their strong bias toward high-abundance proteins and major PTS1 tripeptides. Low-abundance PTS1 proteins, which are derived from weakly expressed genes and occur at very low concentrations in peroxisomes, have only been identified recently, mainly by high-sensitivity proteome analyses of plant peroxisomes (Reumann et al., 2007, 2009; Eubel et al., 2008). Low-abundance PTS1 proteins were noticed to often carry noncanonical PTS1s. Due to this underrepresentation, or even lack, of low-abundance PTS1 proteins in previous data sets and because of their employment of tripeptide-based selection filters, previous PTS1 protein prediction models were not designed to infer novel PTS1 tripeptides or predict low-abundance proteins (Emanuelsson et al., 2003; Neuberger et al., 2003b; Bodén and Hawkins, 2005; Hawkins et al., 2007).

By taking advantage of the large number of EST collections that are available for diverse plant species, we previously generated a data set of 400 PTS1 sequences, leading to the definition of 20 plant PTS1 tripeptides (Reumann, 2004). Six additional PTS1 tripeptides were identified by proteomics-based protein identification in combination with subcellular targeting analysis (SSL>, SSI>, ASL>, SHL>, SKV>, and SLM>; Goepfert et al., 2006; Reumann et al., 2007, 2009; Ma and Reumann, 2008). Including AKI> of *Arabidopsis thaliana*, monodehydroascorbate reductase 1 (MDAR1; Lisenbee et al., 2005) and SRY> of NAD kinase 3 (NADK3; Waller et al., 2010), 28 functional PTS1 tripeptides and 16 position-specific residues ([SAPC] [RKNMSLH] [LMIVY]>) have now been identified for plants. In vivo data suggested that a few additional tripeptides are also functional PTS1s (Mullen et al., 1997) but non-native upstream domains had been used in this study, and plant peroxisomal proteins carrying these tripeptides have not been reported.

The current challenges in PTS1 protein prediction in general, and for plants in particular, are summarized as follows. First, can proteins carrying noncanonical PTS1 tripeptides be correctly predicted? Second, might new prediction methods correctly reveal novel PTS1 tripeptides and residues? Third, can the dependency of PTS1 tripeptides on target-enhancing upstream patterns be inferred from the prediction models?

To increase the number of known plant PTS1 proteins, in general, and of low-abundance proteins in particular, we developed proteomic methods for *Arabidopsis* leaf peroxisomes (Reumann et al., 2007). More than 90 putative novel proteins of peroxisomes, including many low-abundance and regulatory proteins, were thereby identified (Reumann et al., 2007, 2009). By in vivo targeting analysis and PTS identification, a dozen novel *Arabidopsis* PTS1 proteins have been established by our group. These are supplemented by additional proteins identified by the plant peroxisome community with major contributions by the *Arabidopsis* 2010 peroxisome project (www.peroxisome.msu.edu; Ma et al., 2006; Reumann et al., 2007, 2009; Eubel et al., 2008; Moschou et al., 2008; Babujee et al., 2010; Quan et al., 2010; reviewed in Kaur et al., 2009; Reumann, 2011). Many low-abundance proteins carry novel, noncanonical PTS1 tripeptides, further supporting the idea that identification and modeling of low-abundance PTS1 proteins and their targeting signals are prerequisites for the development of prediction tools for low-abundance proteins.

In this study, we generated a large data set of more than 2500 homologous plant sequences, primarily from EST databases, from 60 known *Arabidopsis* PTS1 proteins and developed two prediction methods for plant PTS1 proteins. Both prediction methods showed high accuracy on example sequences and were able to correctly infer novel PTS1 tripeptides, even including novel residues. In combination with large-scale in vivo subcellular targeting analyses, we established 23 newly predicted PTS1 tripeptides for plants and identified several previously unknown *Arabidopsis* PTS1 proteins. Our prediction methods were thereby proven to be suitable for the prediction of plant peroxisomal PTS1 proteins from genomic sequences, including low-abundance and noncanonical PTS1 proteins.

RESULTS

Data Set Generation of PTS1 Protein Example Sequences

First, all known *Arabidopsis* PTS1 proteins (60) were used to identify putatively orthologous full-length cDNAs or predicted protein sequences from other plant species in the nonredundant protein database of GenBank at the National Center for Biotechnology Information. Second, the *Arabidopsis* proteins were tested for their suitability to retrieve putatively orthologous C-terminal sequences from the public database of ESTs, as

described previously (Reumann, 2004). Briefly, plant ESTs that shared the highest sequence similarity with *Arabidopsis* PTS1 proteins but not with *Arabidopsis* paralogs were identified based on sequence similarity above a predefined protein-specific threshold and retrieved irrespective of the identity of their C-terminal tripeptides (see Supplemental Methods online). While more than 90 putatively orthologous sequences were identified for some *Arabidopsis* PTS1 proteins (e.g., ACX1, AGT, MFP2, and SCP2), only a few or none could be detected for other PTS1 proteins (e.g., MCD, OPCL1, UP8, and CSD3; see Supplemental Data Sets 1A and 1B online).

In total, 2562 example sequences of plant PTS1 homologs were retrieved, which were derived from ~260 different plant species. Most sequences originated from dicotyledons (69%), followed by monocotyledons (25%) and other magnoliophyta (e.g., coniferophyta; see Supplemental Data Set 1 online). The majority of sequences (87.2%) were derived from ESTs, demonstrating that ESTs are a major resource for example sequences of plant PTS1 proteins. Because the PTS1 tripeptide is generally the major determinant for peroxisome targeting (see below), sequences with erroneous C-terminal tripeptides would significantly reduce the quality of the data set. Therefore, we separated the data set into three subsets based on the number of sequences that shared the same C-terminal tripeptide. The first, most reliable data subset comprised 96% (2458 sequences) of the example sequences; each of the C-terminal tripeptides was represented by ≥3 sequences. Sequences with tripeptides that were restricted to one or two example sequences were grouped as uncertain sequences in data subsets 2 (26 sequences) and 3 (78 sequences), respectively (Figure 1A; see Supplemental Data Set 1 online).

Forty-two C-terminal tripeptides were identified in a significant number of sequences (≥3, data subset 1) and expected to represent functional PTS1 tripeptides with high probability. Sixteen of these tripeptides had not been proposed to function as targeting signals by previous studies (Table 1). Those tripeptides that had previously been defined as major PTS1 tripeptides based on their abundance in example sequences (Reumann, 2004) generally remained the most abundant and were, in total, present in 85% of the data set sequences. The newly deduced PTS1 tripeptides were each represented by low numbers of sequences in the study sample (see Supplemental Figure 1A online). Likewise, the abundance of position-specific tripeptide residues differed considerably between well-established and newly identified tripeptide residues (see Supplemental Figure 1B online). Sequences upstream of the PTS1 tripeptide are, on average, enriched in Pro, basic residues, and Ser in a position-specific manner (see Supplemental Figure 1C online).

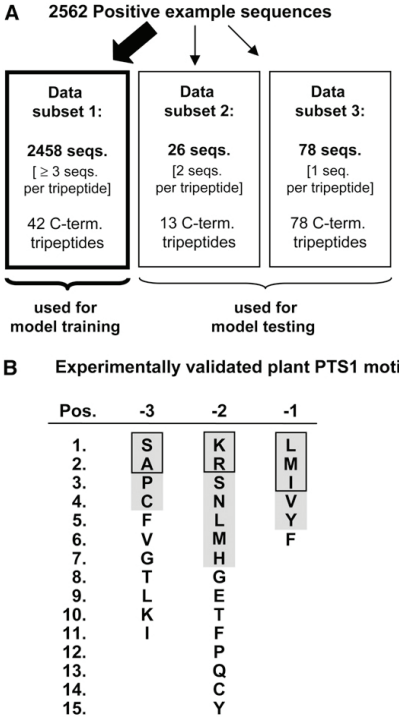


Figure 1. Categorization of Plant PTS1 Protein Example Sequences and Summary of Experimentally Validated Amino Acid Residues Forming the Plant PTS1 Motif.

The 2562 positive example sequences were split into three data subsets according to the number of sequences with the same C-terminal tripeptide. Data set 1, containing 2458 sequences and 42 different C-terminal tripeptides, each represented by ≥3 sequences, was used for training of the prediction models, while data sets 2 and 3 contained unseen sequences and C-terminal tripeptides and were used for model testing. Tripeptide residues previously reported to be present in plant PTS1 tripeptides are shaded in gray. According to experimental data and PWM predictions, at least two of the seven high-abundance residues of high targeting strength ([SA][KR][LMI]>, boxed; see Supplemental Figure 1B online) must be combined with one low-abundance residue to yield functional plant PTS1 tripeptides (x[KR][LMI]>, [SA]y[LMI]>, and [SA][KR]z>).

In Vivo Validations of PTS1 Tripeptides Identified from the Example Data Set

We first investigated whether plant sequences terminating with PTS1 tripeptides that had been deduced from the 2004 data set (Reumann, 2004) but had not yet been experimentally validated could indeed direct a reporter protein to peroxisomes. The PTS1s that we tested included SML>, SNM>, SSM>, SKV>, SRV>, ANL>, and CKL> (Table 1). For each PTS1 tripeptide, one representative example sequence was chosen. The investigated sequences were derived from different enzymes (e.g., sulfite oxidase [SOX] and acyl-activating enzyme isoform 7 [AAE7]) and different plant species (e.g., SSM>, SOX, *Lactuca serriola*; CKL>, AAE7, *Gnetum gnemon*; see Supplemental Table 1 online). The proposed peroxisome targeting domains, comprising the C-terminal decapeptide of the translated ESTs, were attached to a reporter protein, enhanced yellow fluorescent protein (EYFP), and their cDNAs were transiently expressed from the cauliflower mosaic virus 35S promoter in onion epidermal cells

Table 1. Plant PTS1 Tripeptides Deduced from Positive Example Data Sets and/or Predicted by Discriminative Prediction Models and Their Experimental in Vivo Validation

	Plant PTS1 Tripeptides	
Data Set	Newly Predicted	Experimentally Validated in This Study
Data Set-2004	Eight PTS1s and one PTS1 residue: SML>, SNM>, SSM>, SRV>, ANL>, PRM>, **C**KL>, **C**RL>	Six PTS1s and one PTS1 residue: SML>, SNM>, SSM>, SRV>, ANL>, **C**KL>
Data Subset 1-2011	16 PTS1s and seven PTS1 residues: SLL>, SHI>, SNI>, S**G**L>, S**E**L>, S**T**L>, SR**F**>, ALL>, AKM>, CKI>, CRM>, **F**KL>, **F**RL>, **V**KL, **V**RL>, **G**RL>	11 PTS1s and seven PTS1 residues: SLL>, SHI>, S**G**L>, S**E**L>, S**T**L>, SR**F**>, ALL>, CKI>, **F**KL>, **V**KL, **G**RL>
Data Subset 2/3-2011	10 PTS1s and six PTS1 residues: STI>, SGI>, S**F**M>, S**P**L>, S**Q**L>, SEM>, PKI>, **T**RL>, **R**KL>, **L**KL>	Seven PTS1s and five PTS1 residues: STI>, S**F**M>, S**P**L>, S**Q**L>, PKI>, **T**RL>, **L**KL>
Arabidopsis Proteins	Seven PTS1s (plus others) and six PTS1 residues: (SR**Y**>)[1], S**C**L>, S**Y**M>, S**I**L>, S**W**L>, AHL>, **I**KL>, **K**RL>	Five PTS1s and four PTS1 residues: (SR**Y**>)[1], S**C**L>, S**Y**M>, AHL>, **I**KL>, **K**RL>

Newly predicted PTS1 tripeptide residues are underlined and printed bold. With respect to Data Set-2004 (Reumann, 2004), only those tripeptides and residues are indicated that had not been experimentally validated in the meantime. The novel PTS1 tripeptide, SRY>[1], had been identified independently by Waller et al. (2010). Three additional decapeptides investigated in this study represented putative (and validated) non-PTS1 sequences (LCR>, LNL>, and APN>) and are not listed (see Supplemental Tables 1 and 5 online).

that had been biolistically transformed (Fulda et al., 2002). While EYFP alone localized to the cytosol and nucleus, the reporter protein constructs extended by decapeptides terminating with SML>, SNM>, SSM>, ANL>, and CKL> were all observed in punctuate subcellular structures that generally moved quickly along cytoplasmic strands (Figures 2A to 2D, 2F, and 2G). Likewise, the sequence terminating with SKV> targeted EYFP to subcellular organelles, as demonstrated previously for His triad family protein 1 (HIT1; Figure 2E; Reumann et al., 2009). As shown for one representative construct (CKL>), the EYFP-labeled organelles coincided with the cyan fluorescent protein (CFP)-labeled peroxisomes (gMDH-CFP; Fulda et al., 2002), demonstrating that the yellow fluorescent organelles are identical with peroxisomes (Figure 2G).

Peroxisome targeting of EYFP by the chosen SRV> decapeptide of the acyl-CoA oxidase 4 homolog of *Zinnia elegans* could not be resolved under standard conditions (see Supplemental Figure 2A1 online) but required extended expression times (Figure 2H). Under standard conditions of gene expression and protein import into peroxisomes (~18 to 24 h room temperature), the time period of detectable subcellular targeting is limited by the disappearance of cellular reporter protein fluorescence ~24 h after transformation. Vanishing of fluorescence is most likely caused by in vivo degradation of plasmid and EYFP fusion proteins. Consistent with our hypothesis that the process of EYFP degradation is more temperature dependent than protein import into peroxisomes, tissue incubation at reduced temperature (~10°C) significantly extended the time period of observable fluorescence to more than 1 week and made the detection of weak peroxisome targeting possible for several constructs, including the above-mentioned SRV>(1) EST (Figure 2H). The specificity of PTS1 protein import into peroxisomes was verified by EYFP alone and five nonperoxisomal constructs (e.g., LCR> and LNL>; Figure 2A, Ac-Ag), all of which remained cytosolic under the same conditions.

To further confirm SRV> as a plant peroxisomal PTS1, we chose two additional sequences. Indeed, both decapeptides of AGT homologs targeted EYFP to peroxisomes as well, for example, the second sequence [7aa-SRV(2), *Populus trichocarpa* × *Populus deltoides*] with low and the third [7aa-SRV(3), *Pinus taeda*] with high efficiency (Figures 2I and 2J; see Supplemental Figures 2B1 and 2B2 online). The differential peroxisome targeting efficiency of different decapeptides carrying the same noncanonical PTS1 tripeptides indicates the strong dependence of noncanonical PTS1 tripeptides on the presence and strength of targeting enhancing patterns located upstream of the PTS1 tripeptide to cause peroxisome targeting (see also below).

Taken together, six previously predicted tripeptides (Reumann, 2004) were thereby established, in the context of the 10–amino acid targeting domain of native PTS1 proteins, as functional plant PTS1 tripeptides. Additionally, Cys was experimentally validated as a PTS1 tripeptide residue at position −3, as indicated previously (Table 1; Reumann, 2004). These results confirmed the quality of the previous and present data sets of PTS1 protein example sequences and the reliability of our approach in identifying functional plant PTS1 tripeptides from homologous ESTs (Reumann, 2004).

We next set out to experimentally validate the 16 novel PTS1 tripeptides that had been deduced from the present example sequences (example data set 1, Figure 1). Seven tripeptides represented previously unknown combinations of known tripeptide residues, while nine PTS1 tripeptides contained seven residues that had not previously been shown to exist in the plant PTS1 motif (Table 1, Figure 1B). Indeed, the four representative decapeptides that we investigated terminating with novel combinations of known PTS1 residues, including SHI>, SLL>, ALL>, and CKI> (Table 1; see Supplemental Table 1 online), all targeted EYFP to small subcellular structures under standard expression conditions (Figures 2K, 2L, 2N, and 2O). The identity of the fluorescent structures with peroxisomes was verified representatively for two constructs (ALL> and CKI>; Figures 2N and 2O).

Regarding the reporter protein constructs extended by decapeptides with novel tripeptide residues, all proteins targeted to peroxisomes as well, although some did so with low efficiency

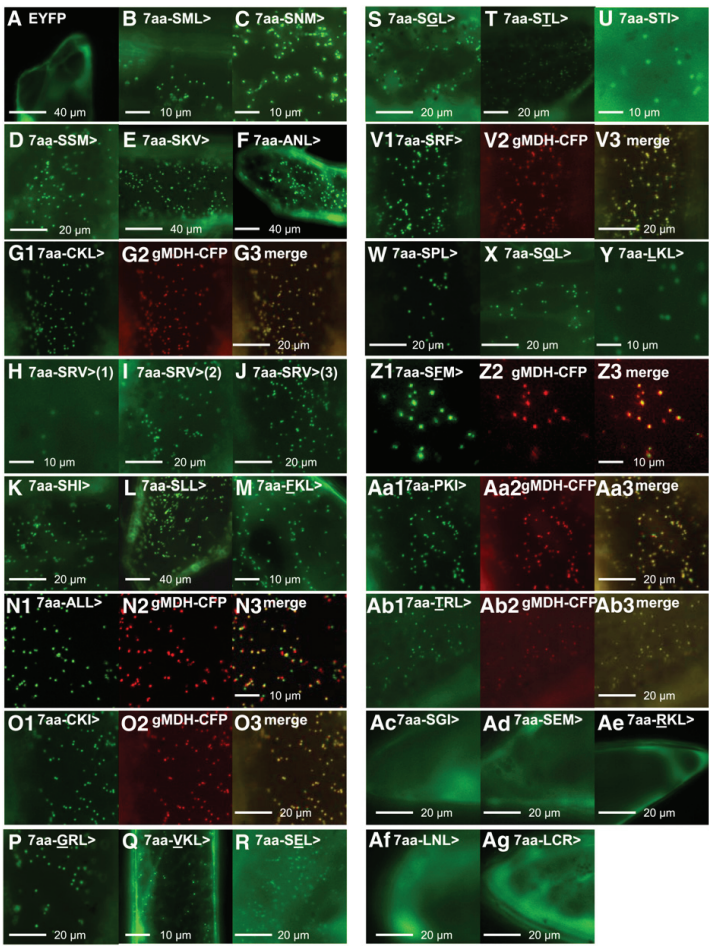


Figure 2. Experimental Validation of Example Sequences by in Vivo Subcellular Targeting Analysis.

Onion epidermal cells were transformed biolistically with EYFP fusion constructs that were C-terminally extended by the C-terminal decapeptides of plant PTS1 proteins serving as example sequences. Subcellular targeting was analyzed by fluorescence microscopy after ~18 h expression at room temperature only (**[B]**, **[C]**, **[E]** to **[G]**, **[J]** to **[O]**, **[Q]**, **[T]**, **[V]**, **[X]**, **[Z]**, **[Aa]**, and **[Ab]**), at an additional 24 h at ~10°C (**[A]** and **[Ac]** to **[Ag]**), or at an additional 5 to 6 d at ~10°C (**[D]**, **[H]**, **[I]**, **[P]**, **[R]**, **[S]**, **[U]**, **[W]**, and **[Y]**). Cytosolic constructs, for which subcellular targeting data are shown after short-term expression times, were reproducibly confirmed as cytosolic also after long-term expression. Novel amino acid residues of PTS1 tripeptides are underlined. In double transformants, peroxisomes were labeled with CFP, and cyan fluorescence was converted to red for image overlay (**[G]**, **[N]**, **[O]**, **[V]**, **[Z]**, **[Aa]**, and **[Ab]**). To document the efficiency of peroxisome targeting, EYFP images of single transformants were not modified for brightness or contrast. The sequences that terminated with LNL> and LCR> were included as putative non-PTS1 sequences (**[Af]** and **[Ag]**). Comparative subcellular targeting results obtained under different expression conditions are shown in Supplemental Figure 2 online. For sequence details, see Supplemental Tables 1 and 6 online.

(e.g., FKL> and VKL>; Figures 2M and 2Q). Extended expression times at low temperature improved peroxisome targeting for some (e.g., SGL>, Figure 2S and Supplemental Figure 2G1/2 online; SEL>, Figure 2R and Supplemental Figure 2F1/2 online; STL>, Figure 2T) but not all constructs (e.g., FKL>, Figure 2M and Supplemental Figure 2C1/2 online; GRL>, Figure 2P and Supplemental Figure 2D1/2 online; VKL>, Figure 2Q and Supplemental Figure 2E1/2 online; STI>, Figure 2U and Supplemental Figure 2H1/2 online). Peroxisome targeting mediated by SEL>, which atypically carried the acidic residue, Glu, at position −2,

was particularly weak and could only be resolved after extended expression times. Taken together, the decapeptides comprising novel residues (underlined) in the predicted PTS1 tripeptides, including FKL>, GRL>, and VKL> (with Phe, Gly, or Val at position −3), SEL>, SGL>, and STL (with Glu, Gly, or Thr at position −2), and SRF> (with Phe at position −1), all targeted EYFP to punctuate subcellular structures (Figures 2M, 2P to 2T, and 2V). Coincidence of the EYFP-labeled organelles with peroxisomes was representatively verified for SRF> (Figure 2V).

In summary, all 11 newly identified PTS1 tripeptides that were subjected to experimental analysis were confirmed as functional PTS1s. The experimental data that have been presented so far have increased the number of experimentally verified plant PTS1 tripeptides by 17 and established seven additional residues within the plant PTS motif ([FVG][GET]F>, Figure 1B). Seven additional closely related tripeptides, which were also represented by ≥3 example sequences but not investigated experimentally, are likely to also function as plant PTS1 tripeptides (SNI>, AKM>, PRM>, CRL>, CRM>, FRL>, and VRL>; Table 1).

Development of Two Discriminative Prediction Methods for Plant PTS1 Proteins

We concluded from the high experimental verification rate of newly predicted PTS1 tripeptides (see above) that data subset 1 (Figure 1A) was a reliable set of positive example sequences that was suitable for the development of discriminative PTS1 protein prediction algorithms. A data set of 21,028 negative example sequences from spermatophyta (seed plants) was additionally generated (see Supplemental Methods online). For both types of example sequences, a maximum of 15 C-terminal amino acid residues was considered. Two different discriminative prediction methods were applied: (1) position-specific weight matrices (PWMs) and (2) residue interdependence (RI) models. While PWM models are trained using only position-specific amino acid abundances in the example sequences, RI models are able to consider possible dependencies between amino acid residues, for instance, between the PTS1 tripeptide and upstream residues. For learning of discriminative models we used so-called regularized least squares classifiers (see Supplemental Methods online; Rifkin et al., 2003). In contrast with the methods used in previous PTS1 protein prediction studies (Emanuelsson et al., 2003; Neuberger et al., 2003b, 2003a; Bodén and Hawkins, 2005, Hawkins et al., 2007), these classifiers offer three major advantages. First, they provide interpretable discriminative features in terms of important amino acid residues or residue interdependencies. Second, these classifiers allow fast prediction of potential PTS1 proteins in complete genomes and whole databases. Third, our prediction models do not involve any preselection filters for PTS1 tripeptides, which had been applied in previous PTS1 prediction tools (Emanuelsson et al., 2003; Bodén and Hawkins, 2005; Hawkins et al., 2007). PTS1 tripeptide filters restrict the prediction of PTS1 proteins to those carrying known PTS1 tripeptides (Bodén and Hawkins, 2005; Hawkins et al., 2007) or residues (Emanuelsson et al., 2003). Our prediction models could potentially predict proteins with previously unidentified PTS1 tripeptides as peroxisomal and, moreover, infer novel PTS1 tripeptide residues.

The prediction sensitivity (i.e., the rate at which positive examples are correctly predicted as peroxisomal) was high for both prediction models. If the PTS1 tripeptide alone was considered, 95% (PWM) of the positive example sequences were already correctly predicted as peroxisome targeted (0.95 sensitivity; Figure 3), confirming that the PTS1 tripeptide is generally the major discriminative determinant for peroxisome targeting. With increasing size of the PTS1 domain, the prediction sensitivity further increased. Maximum sensitivity was achieved by taking into consideration the 14 (PWM model, 0.981) or 15 C-terminal amino acid residues (RI model, 0.996; see Supplemental Table 2 online). Hereby, the order in which the upstream residue positions were added to the prediction model was not important (i.e., the prediction performance depends on the number of residues instead of the distance of the residues from the C terminus) (see Supplemental Table 3 and Supplemental Methods online for details).

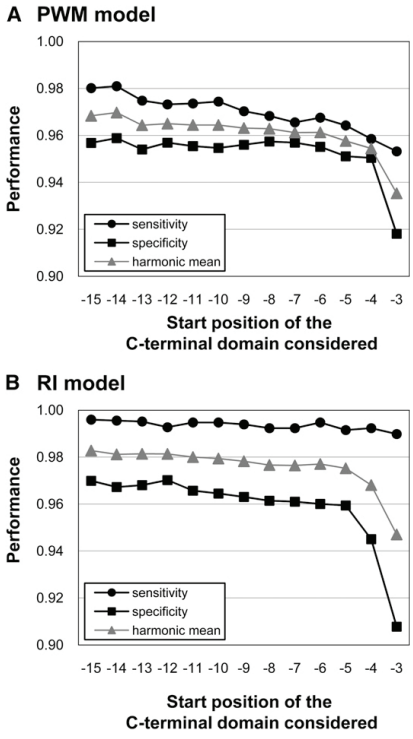


Figure 3. Performance Analysis of the PWM and RI Prediction Models on Example PTS1 Protein Sequences.

The *x* axis indicates the start position of the C-terminal PTS1 domain that was considered for performance analysis and extends to the extreme C termini of the PTS1 proteins. For the definition of sensitivity, specificity, and harmonic mean, see Supplemental Methods online.

The prediction specificity, which indicates how many positively predicted proteins are indeed peroxisomal, was also high for both prediction models (0.959 for the PWM and 0.970 for the RI model). The harmonic mean of prediction sensitivity and specificity was optimal for the C-terminal 14 (PWM model, 0.970) and 15 amino acid residues (RI) and slightly higher for the RI model (0.983; Figure 3; see Supplemental Table 2 online). To check whether keeping highly similar sequences influences the prediction performance during cross-validation, we also evaluated our models on a version of the data set that had been reduced to 50–amino acid sequences sharing ≥90% sequence similarity (for details, see Supplemental Methods online). No substantial decline of the prediction performance was observed (see Supplemental Table 3 online).

Because of their high performance, both the PWM and RI models were applied to the positive and negative example data sets and provided two independent prediction scores for each example sequence. The prediction threshold, which is the score that corresponds to a 50% probability of peroxisome targeting according to the model, was calculated as 0.412 (PWM model) and 0.219 (RI model). To facilitate interpretation of the absolute prediction scores, model-specific posterior probabilities were calculated, which quantify the probability for peroxisome targeting (see Supplemental Methods online). These probability values range from zero (0% probability) to one (100%), with 0.5 corresponding to the prediction threshold that assigns to the sequences with this value a 50% probability for peroxisome targeting. The dependency of the posterior probabilities on the prediction score for both models is illustrated in Supplemental Figure 3 online. The steepness of the graph is higher for the RI model, which is a consequence of its higher model complexity.

Only 2.0% of the positive and 0.4% of the negative examples were predicted incorrectly by the PWM model. The incorrectly predicted negative example sequences likely include both peroxisomal proteins that are as yet unknown/unannotated to be peroxisome targeted and obviously false predictions. The RI model correctly predicted all of the positive example sequences and 99.9% of the negative example sequences (see Supplemental Data Set 1B online). In summary, the prediction accuracy of both models was high. Despite the absence of any selection filter for known PTS1 tripeptides, both prediction models maintained high prediction specificity. The RI model performed slightly better on example sequences compared with the PWM model. Moreover, the discriminative models used in this study are computationally very efficient as predictors of novel peroxisomal protein sequences: the prediction of 21,028 (negative) example sequences using 15 C-terminal residues took 0.34 s for the PWM and 0.37 s for the RI model on a 2.83-GHz Xeon processor (see Supplemental Table 2 online). This low evaluation time (<0.02 ms/sequence) makes it possible to scan whole genomes or even complete databases in a few seconds.

Out of the 20 constructs that carry noncanonical tripeptides, all of which have been experimentally validated as peroxisomal thus far, 20 and 14 were correctly predicted by the RI and PWM models, respectively. The PWM model predicted the other six peroxisomal proteins as cytosolic [SRF>, SGL>, SRV>(1), SKV>, CKI>, and SEL>; see Supplemental Table 1 online]. The data further confirmed that the RI model performed better on example sequences compared with the PWM model (see Supplemental Table 3 online).

Experimental Model Validation on Example Sequences Carrying Unseen Tripeptides

In general, the data sets that have been used in previous studies (Picard and Cook, 1984; Emanuelsson et al., 2003; Bodén and Hawkins, 2005; Hawkins et al., 2007) and in the first part of our article (data subset 1, Figure 1A) are biased toward canonical PTS1 tripeptides. To test our algorithms with respect to their ability to predict unseen PTS1 patterns, we applied them to sequences (and C-terminal tripeptides) that had been excluded completely from model training and validation (i.e., data subsets 2 and 3) (Figure 1A; see Supplemental Data Sets 1A and 1B and Supplemental Table 1 online). Representative example sequences were selected for experimental verification based on their ability to introduce novel residues into the plant PTS1 motif and on their PWM and RI model-based prediction scores with the goal of systematically covering the score ranges below the thresholds. In this manner, 12 additional example sequences were chosen for experimental validation, including two putative non-PTS1 sequences (LCR> and LNL>) that deviated from the emerging PTS1 tripeptide pattern (x[KR][LMI]>, [SA]y[LMI]>, and [SA][KR]z>; Figure 1B; see Supplemental Table 1 online and Discussion).

The C-terminal decapeptides of seven sequences indeed targeted EYFP to small subcellular organelles, although with different efficiency (STI>, SPL>, SQL>, SFM>, PKI>, TRL>, and LKL>; Figures 2U and 2W to 2Ab; see Supplemental Table 1 online). The specificity of PTS1 protein import into peroxisomes was further confirmed by the two suspected non-PTS1 sequences (LCR> and LNL>) that remained cytosolic under the same conditions (Figures 2Af and 2Ag). The identity of the fluorescent organelles as peroxisomes was verified by three representative decapeptides (SFM>, PKI>, and TRL>; Figures 2Z to 2Ab). These in vivo analyses identified seven additional novel PTS1 tripeptides (STI>, SPL>, SQL>, SFM>, PKI>, TRL>, and LKL>) and added five novel residues, namely, Thr and Leu (position −3) and Pro, Phe, and Gln (position −2) to the plant PTS1 tripeptide motif ([TL][PFQ]z>). Three other EYFP constructs (SGI>, SEM>, and RKL>) remained cytosolic, further confirming the specificity of peroxisome import (Figures 2Ac to 2Ae; see Supplemental Table 1 online). The results supported our initial assumption that the ESTs of these two uncertain data subsets are less reliable and may contain erroneous amino acid residues either in the C-terminal tripeptide or the upstream region that prohibit peroxisome targeting (see Discussion).

Assessing the prediction accuracy of the models for these 12 sequences, four to five cytosolic sequences were confirmed to have been correctly predicted, while six to seven peroxisome-targeted sequences had been scored slightly below the threshold by both models. Importantly, one verified PTS1 domain (SQL>) had correctly been predicted by the PWM model as peroxisomal, although SQL> sequences and sequences with Q at position −2 in general had been completely absent from the training data set. Likewise, another novel PTS1 tripeptide, SFM>, was predicted as peroxisomal with relatively high posterior

probability (0.40) but was slightly below the threshold (see Supplemental Table 1 online). Three major conclusions were drawn from the predictions and experimental validations of sequences carrying unseen PTS1 tripeptides: (1) both models tend to score peroxisomal sequences with novel PTS1 tripeptides below the threshold and can thus be considered as conservative predictors with respect to unseen PTS1 patterns; (2) despite its slightly inferior performance on training data, the PWM model performed better in pattern abstraction from training to unseen sequences compared with the RI model; and (3) the PWM model is able to correctly predict peroxisomal proteins with previously unseen PTS1 tripeptides (SQL>), which even included one novel tripeptide residue (Q, position −2).

Differential Dependence of PTS1 Tripeptides on Targeting-Enhancing Upstream Patterns

Apart from the reported role of basic residues in enhancing protein targeting to peroxisomes by the PTS1 pathway (Distel et al., 1992; Kragler et al., 1998; Bongcam et al., 2000; Brocard and Hartig, 2006; Ma and Reumann, 2008), little information is available on the identity of such patterns and their quantitative effect on peroxisome targeting. To investigate the predicted influence of the upstream region on peroxisome import, we analyzed the most discriminative weights of both models. The positive (negative) discriminative weights reflect features of the upstream region that are overrepresented (underrepresented) in the positive example sequences. The PWM model allows inference of the importance of certain features in terms of the position-specific absence or presence of a particular residue. Our learned PWM model indicated that Trp (W, positions −14 and −13), Pro (P, positions −5, −7, and −10), and basic residues (R, positions −4 and −6; H, position −4) are helpful in directing proteins into peroxisomes. On the other hand, the large negative weights for W at position −6 and Tyr (Y) at position −11 indicate their negative effect on peroxisome targeting (see Supplemental Table 4 online). The RI-based model revealed possible interdependencies of residues at particular positions and indicated, for instance, a positive influence of P (positions −5 and −7) and basic residues (K, positions −4, −7, and −8; R, pos. −4) in the upstream region in combination with the tripeptide residues, S (position −3) and L (position −1). By contrast, the RI model showed large negative weights for dimensions associated with the occurrence of the residues G, D, and E (position −4) and L (positions −14 and −13), suggesting a pronounced prohibitive effect of these residues on peroxisome targeting (see Supplemental Table 4 online).

To address whether the models predicted the PTS1 tripeptides to differ in strength and dependency of targeting-enhancing upstream patterns, we computed the prediction scores for the 42 data set–deduced PTS1 tripeptides (see Supplemental Figure 1A online) in the context of all possible combinations of a maximum number of upstream residues (i.e., upstream hexapeptides, for example, for 42*64,000,000 nonapeptides). For most major PTS1 tripeptides (e.g., SKL> and ARL>), the PWM model predicted >95% of the nonapeptides as peroxisome targeted, indicating that major PTS1 tripeptides are strong and mediate peroxisome targeting nearly independently of the upstream domain (see Supplemental Figure 4A online). The corresponding RI model-based predictions showed the same tendency but at a lower rate (70 to 90%), indicating a higher stringency of PTS1 protein prediction. By contrast, for most minor and noncanonical PTS1s (e.g., SRV>, SHI>, ALL>, and GRL>; see Supplemental Figures 1 and 4 online), both models predicted <10% of the nonapeptide combinations as peroxisome targeted, assigning to these PTS1 tripeptides weak targeting strengths and strong dependencies on specific targeting-enhancing upstream patterns for functional activity. Moreover, single amino acid residue exchanges in PTS1 tripeptides are predicted to drastically reduce the targeting strength of the tripeptide itself (e.g., PWM: SR [LMI]>, 85 to 99% nonapeptides peroxisomal; SRV>, 0.9%; see Supplemental Figure 4A online). In summary, and consistent with previous experimental indications (see above), the two models quantitatively assign high targeting strengths to major PTS1 tripeptides and low strengths and pronounced dependencies on targeting enhancing upstream patterns to noncanonical PTS1s.

To investigate the variability of targeting-enhancing patterns, we analyzed the position-specific amino acid composition of the upstream hexapeptide of peroxisome-predicted nonapeptides. We representatively selected three noncanonical PTS1 tripeptides associated with comparatively few peroxisome-predicted nonapeptide combinations, ALL>, SKV>, and SRF>, for this analysis. While the ALL-containing nonapeptides predicted to be peroxisome targeted are, on average, enriched for Arg (positions −4 and −6) and, to a minor extent, for His (positions −7 and −8), the corresponding SRF> and SKV> nonapeptides are highly enriched for Pro (position −7; see Supplemental Figures 4B to 4D online). The data further supported the hypothesis that basic residues and P are major targeting-enhancing residues in plant peroxisomal PTS1 proteins (Reumann, 2004) and indicate that targeting-enhancing patterns are complex and differ among different noncanonical PTS1 tripeptides.

PTS1 Protein Predictions from the *Arabidopsis* Genome and Experimental Validations

We next applied both prediction models to the *Arabidopsis* genome. The TAIR10 database (release November 2010) comprises 35,385 proteins (or gene models) that include transcriptional and translational variants derived from 27,416 gene loci. Prediction scores and posterior probabilities were calculated for all *Arabidopsis* gene models using the PWM and RI prediction methods, thereby providing a hierarchical list of all *Arabidopsis* gene models according to their peroxisome targeting probabilities (see Supplemental Figure 5 and Supplemental Data Set 2 online). In total, 392 *Arabidopsis* proteins (1.1% of the genome, 320 loci) were predicted to be PTS1 proteins targeted to peroxisomes (Figure 4). These gene models included 109 gene models (79 gene loci) encoding established plant peroxisomal PTS1 proteins and 12 additional gene models (10 gene loci) that have been associated with plant peroxisomes based on proteomics data only up to now. Approximately 271 gene models (231 gene loci) had not yet been associated with peroxisomes, indicating that up to 70% of *Arabidopsis* PTS1 proteins might have remained unidentified up to now (see Supplemental Data Set 2 online).

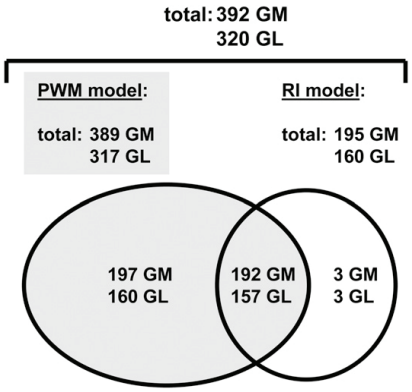


Figure 4. Venn Diagram of PWM- and RI-Model Based PTS1 Protein Predictions for *Arabidopsis*.

The 392 gene models (GM; i.e., transcriptional and translational protein variants) and 320 gene loci (GL; i.e., protein coding genes) are predicted PTS1 proteins by either the PWM or the RI model. Except for three proteins (At1g21770.1, At4g02340.1, and At5g02660.1), the RI model predicted a protein subset of those predicted by the PWM model to be peroxisome-targeted PTS1 proteins. For details on PWM and RI model predictions for the 35,385 *Arabidopsis* gene models (TAIR10, November, 2010; 27,416 loci), see Supplemental Data Set 2 online. The 392 gene models (320 gene loci) include 109 gene models (79 gene loci) encoding established plant peroxisomal PTS1 proteins, 12 gene models (10 gene loci) associated with plant peroxisomes based on proteomics data only, and 271 gene models (231 gene loci) that had not yet been associated with peroxisomes, indicating that up to 70% of *Arabidopsis* PTS1 proteins might have remained unidentified up to now.

The PWM model predicted 389 proteins as peroxisome targeted (see Supplemental Data Set 2 online), while the RI model was more restrictive and predicted 195 PTS1 proteins. Except for three proteins, the PTS1 proteins that were predicted by the RI model represented a subset of those predicted by the PWM model (Figure 4). Five recently established peroxisomal PTS1 proteins were scored below the thresholds (see Supplemental Data Set 2 online).

Consistent with the nonapeptide analysis (see above), both prediction models assigned a differential dependence on targeting-enhancing upstream patterns to PTS1 tripeptides in *Arabidopsis* proteins. Consistent with the general independence of major PTS1 tripeptides on targeting-enhancing upstream patterns, nearly all *Arabidopsis* gene models carrying major known PTS1s were predicted as peroxisomal (e.g., PWM model: SKL>, 52 out of 52 gene models; ARL>, 20/20; PKL>: 13/13). By contrast, for newly identified noncanonical PTS1s, only a few, specific gene models carrying targeting enhancing upstream patterns were predicted as peroxisome targeted (e.g., SKV>, 3/16; SRY>, 1/7; SPL>, 3/15; see Supplemental Data Set 2 online). A few, specific *Arabidopsis* proteins carrying particular noncanonical PTS1s (e.g., SPL> and VKL>) and suitable targeting-enhancing upstream patterns will thus be peroxisome-targeted in vivo, while most SKV> and VKL> proteins lack such targeting-enhancing upstream patterns and will be cytosolic.

Compared with the positive example sequences of data sets 1 to 3 (Figure 1A; see Supplemental Data Set 1 online; see above), the prediction of unknown proteins as PTS1 proteins from genome sequences requires an even more advanced abstraction and inference ability from the models. In this task, the prediction models not only have to deal with C-terminal tripeptides that had been absent from the training data set, but also with proteins that lack any sequence homology to those used for model training. We therefore validated the genomic PTS1 protein predictions in detail and subjected another set of representative proteins to in vivo subcellular targeting analysis. Because major PTS1 tripeptides mediate peroxisome targeting largely independently of their upstream domains (see above), the C-terminal decapeptides of unknown *Arabidopsis* proteins with major PTS1 tripeptides are unlikely not to target a reporter protein to peroxisomes. Consequently, these proteins were considered to be less suitable for critical testing of these predictions. Instead, we largely focused on the most challenging predictions (i.e., proteins carrying noncanonical or previously undiscovered PTS1 tripeptides). We chose 20 additional *Arabidopsis* proteins with the goal of verifying the predictions thoroughly, discovering novel plant PTS1 tripeptides and identifying novel low-abundance proteins of important physiological function (see Supplemental Table 5 online). Both C-terminal decapeptides and full-length protein fusions with EYFP were analyzed.

We first investigated subcellular targeting of EYFP extended C-terminally by predicted PTS1 domains of *Arabidopsis* proteins. Among the 15 reporter constructs tested, 10 were targeted to punctuate subcellular structures. Colocalization of these structures with peroxisomes was confirmed using four representative constructs (Figures 5A, 5H, 5L, and 5M; see Supplemental Table 5 online). The *Arabidopsis* proteins that were validated to carry functional PTS1 domains included one unknown protein (UP9, SCL>), a 1-aminocyclopropane-1-carboxylate synthase like pseudogene [ACS3, SPL>(2)], a Tudor superfamily protein (Tudor, KRL>), short-chain dehydrogenase/reductase isoform c (SDRc, SYM>), a GTP binding protein (SPK1, SEL>), a PHD finger family protein (PHD, SRY>), a lecithin:cholesterol acyltransferase family protein (LACT, IKL>), calcium-dependent protein kinase isoform 1 (CPK1, LKL>), and purple acid phosphatase 7 (PAP7, AHL>; Figures 5A, 5C, 5E, 5F, 5H, 5I, and 5K to 5N). Moreover, our elevated detection sensitivity allowed the visualization of peroxisome targeting achieved by the C-terminal domain of a protein kinase, which had previously remained undetected (PK1, Figure 5P; Ma and Reumann, 2008).

The prediction algorithms thereby allowed, out of 35,385 gene models, straightforward identification of 10 additional *Arabidopsis* proteins with functional noncanonical PTS1 domains, most of which carried unknown PTS1 tripeptides. Consistent with the noncanonical nature of the predicted PTS1 tripeptides and largely consistent with the model predictions, the C-terminal domain constructs of five other *Arabidopsis* proteins remained cytosolic [SPL>(1), SWL>, APN>, SIL>, and VKL>; Figures 5B, 5D, 5G, 5J, and 5O; see Supplemental Table 5 online]. Cytosolic targeting of the *Arabidopsis* VKL> protein (CUT1) as opposed to

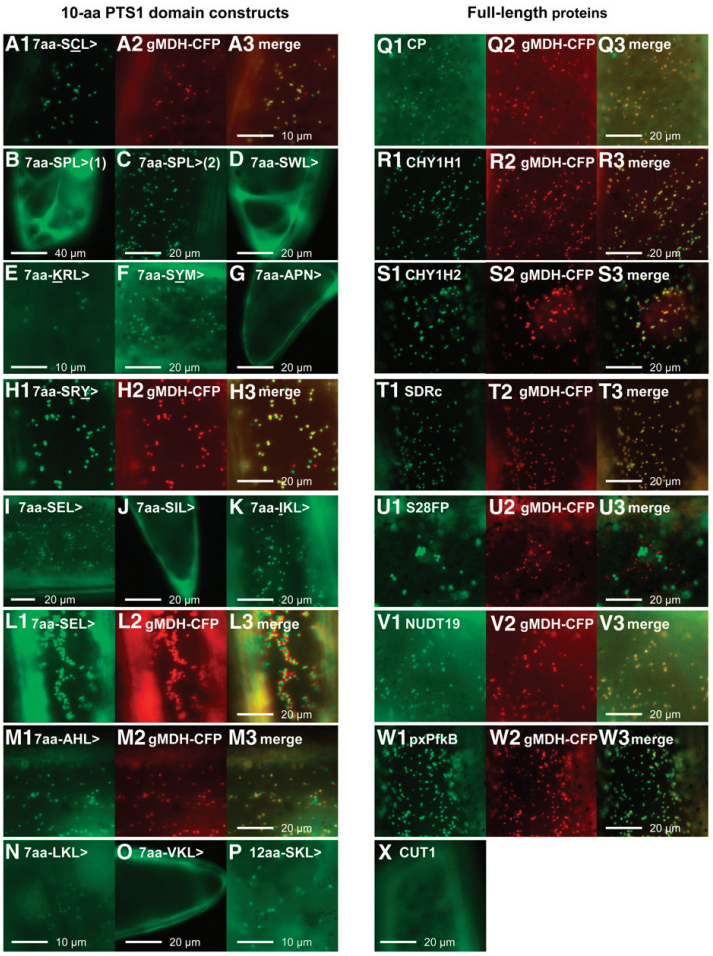


Figure 5. Experimental Validation of *Arabidopsis* Proteins Newly Predicted to Be Located in Peroxisomes by in Vivo Subcellular Targeting Analysis.

Onion epidermal cells were transformed biolistically with EYFP fusion constructs that were either C-terminally extended by the C-terminal decapeptide of representative *Arabidopsis* proteins (or the 15–amino acid peptide for PK1, P) or fused with *Arabidopsis* full-length cDNAs. Novel amino acid residues of newly identified functional PTS1 tripeptides (in addition to those identified in Figure 2) are underlined. Subcellular targeting was analyzed by fluorescence microscopy after ~18 h expression at room temperature only (**[A]** to **[C]**, **[F]**, **[H]**, **[I]**, **[K]**, **[M]**, **[R]** to **[T]**, **[W]**, and **[X]**), at an additional 24 h at ~10°C (**[D]**, **[E]**, **[G]**, **[J]**, **[N]** to **[Q]**, **[U]**, and **[V]**), or at an additional 5 to 6 d of expression at ~10°C **(L)**. Cytosolic constructs, for which subcellular targeting data are shown after short-term expression times, were reproducibly confirmed as cytosolic also after long-term expression. In double transformants, peroxisomes were labeled with CFP, and cyan fluorescence was converted to red for image overlay (**[A]**, **[H]**, **[L]**, **[M]**, and **[Q]** to **[W]**). The predicted PTS1 domains investigated derived from the following proteins: SCL> (UP9), SPL>(1) (FAH), SWL> (RING), KRL> (Tudor), SYM> (SDRc, At3g01980.1/3/4), APN> (SDRc, At3g01980.2), SEL> (SPK1), SRY> (PHD), SIL> (ANK), IKL> (LCAT), LKL> (CPK1), VKL> (CUT1), AHL> (PAP7), and PK1 (SKL>; Ma and Reumann, 2008). The predicted PTS1 tripeptides of the *Arabidopsis* full-length proteins are the following: CP (SKL>), CHY1H1 and CHY1H2 (both AKL>), SDRc (SYM>), S28FP (SSM>), NUDT19 (SSL>), pxPfkB (SML>), and CUT1 (VKL>). To document the efficiency of peroxisome targeting, EYFP images of single transformants were not modified for brightness or contrast. The Arabidopsis Genome Initiative codes of the *Arabidopsis* proteins are listed in Supplemental Table 5 online.

peroxisome targeting of the VKL> example EST (Figure 2Q), both correctly predicted by the PWM model, is explained by the presence of essential targeting enhancing upstream elements in the latter that lack in the former.

Among the 10 *Arabidopsis* proteins verified to carry functional PTS1 domains, eight had been correctly predicted as peroxisomal proteins by the PWM model, supplemented by CPK1 with a prediction score slightly below threshold (0.321, 8% posterior probability), indicating that the prediction accuracy of the PWM model on *Arabidopsis* proteins was particularly high. Except for SEL> and SPL>, all of these validated PTS1 tripeptides (SCL>, SYM>, SRY>, KRL>, and IKL>) had been absent from the training data set, demonstrating that the PWM model was able to correctly predict several novel PTS1 tripeptides. The PWM model could not only infer novel combinations of known position-specific residues, but it could also predict PTS1 tripeptides with novel amino acid residues ([KI][CY]Y>). The RI model inferred the novel PTS1 tripeptides of two *Arabidopsis* proteins correctly (SCL> and SYM>) but seemed too restrictive for the purpose of pattern abstraction.

We finally investigated whether fusions between *Arabidopsis* full-length proteins and the reporter protein were peroxisome localized, which is prerequisite to conclusively identifying novel PTS1 proteins. Out of eight *Arabidopsis* proteins tested, six proteins were confirmed as peroxisome targeted. A Cys protease (SKL>) was targeted to organelles, coincident with CFP-labeled peroxisomes in double transformants (Figure 5Q). The full-length cDNAs of two CHY1 homologs (CHY1H1 and CHY1H2, AKL>) likewise were shown to be located in peroxisomes (Figures 5R and 5S). Short-chain dehydrogenase/reductase isoform c (SDRc), for which three out of four gene models carry the atypical PTS1-related tripeptide, SYM>, also targeted EYFP to peroxisomes (Figure 5T). Alternative in vivo splicing of the cDNA of variant 2 (At3g01980.2, APN>) to other SDRc variants (At3g01980.1/3/4, SYM>) was verified by more detailed peroxisome targeting analysis. While the reporter protein containing the decapeptide terminating with SYM> was targeted to peroxisomes, the construct terminating with APN> remained cytosolic (Figures 5F and 5G; see Supplemental Table 5 online).

The full-length protein of a Ser carboxypeptidase S28 family protein (S28FP, SSM>) directed EYFP to subcellular vesicle-like structures that did not coincide with peroxisomes (Figure 5U). Nudix hydrolase homolog 19 (NUDT19, SSL>) appeared to carry a weak PTS1 domain (Figure 5V). PfkB-type carbohydrate kinase family protein (pxPfkB, SML>) was also verified as a peroxisomal protein (Figure 5W). Only a single full-length protein tested remained cytosolic (CUT1, VKL>; Figure 5X), consistent with both model predictions, the noncanonical nature of its C-terminal tripeptide, and the in vivo data for its C-terminal domain (Figure 5O; see Supplemental Table 5 online).

Taken together, the experimental analyses identified 11 novel *Arabidopsis* proteins carrying noncanonical PTS1 tripeptides. To investigate the significance of the PTS1 protein prediction tools, we analyzed whether these proteins would have been correctly predicted as peroxisomal by other Web tools. However, only four proteins (PTS1 predictor) or even none (PeroxiP) out of 11 newly identified *Arabidopsis* proteins carrying noncanonical PTS1 tripeptides were correctly predicted as peroxisomal by preexisting PTS1 protein prediction tools (see Supplemental Table 5 online), demonstrating the necessity and significance of the new PTS1 protein prediction tools for plant research.

In summary, the in vivo localization data for previously unidentified *Arabidopsis* peroxisomal proteins (1) demonstrated that five additional tripeptides are plant PTS1s (SCL>, SYM>, IKL>, KRL>, and AHL>), (2) added four novel residues to the PTS1 tripeptide motif ([IK][CY]z>), (3) determined that 10 *Arabidopsis* proteins carry functional PTS1 domains, and (4) established six additional *Arabidopsis* proteins as novel peroxisomal proteins. Both prediction models were able to infer novel PTS1 tripeptides, including novel tripeptide residues, with the best performance being evident for the PWM model.

DISCUSSION

Experimental proteome analyses of peroxisomes have recently been reported for model plant species such as *Arabidopsis*, soybean (*Glycine max*), and spinach (*Spinacia oleracea*) (Fukao et al., 2002, 2003; Reumann et al., 2007, 2009; Eubel et al., 2008; Arai et al., 2008a, 2008b; Babujee et al., 2010). Combined with in vivo subcellular targeting analyses, these studies have significantly extended the number of established peroxisomal matrix proteins and broadened our knowledge of peroxisome metabolism (Kaur et al., 2009; Reumann, 2011). Despite their success, these studies are limited in their protein identification abilities by several parameters, for instance, by technological sensitivity and peroxisome purity, and to major plant tissues and organs. Additionally, only a few model plant species are suitable for peroxisome isolation, and the plants must generally be grown under standard rather than environmental or biotic stress conditions, which enhance organelle fragility. These experimental limitations can be best overcome by the development of high-accuracy prediction tools for plant peroxisomal matrix proteins, their application to plant genomes, and relatively straightforward in vivo validations of newly predicted proteins (Reumann, 2011). High-accuracy prediction tools have been lacking for plants up to now. Because ~80% of matrix proteins enter plant peroxisomes by the PTS1 import pathway (Reumann, 2004), prediction algorithms for PTS1 proteins are expected to significantly contribute to defining the plant peroxisomal proteome.

High PTS1 Protein Prediction Sensitivity

High-accuracy prediction models are characterized by both high prediction sensitivity and specificity. The gold standard in bioinformatics to determine these performance parameters is to randomly split data sets of example sequences into different subsets, some of which are used for model training, while a disjoint set is used for testing of the prediction accuracy (see Supplemental Methods online). In this approach, both models yielded high performance values of >98% sensitivity and >96% specificity (Figure 3; see Supplemental Table 2 online).

The prediction sensitivity of a model in detecting plant PTS1 proteins mainly depends on the ability to identify all functional PTS1 tripeptides of Spermatophyta. In this study, novel plant PTS1 tripeptides were identified by two methods: direct

identification from a data set of plant PTS1 sequences and correct inference by prediction models. Careful manual identification of homologous sequences in EST databases allowed the generation of a large data set of PTS1 sequences (87% translated ESTs) from 260 plant species. The size of this data set exceeds that of other metazoan studies, all of which were restricted to protein sequences, by at least eightfold (2500 compared with 90 to 300 sequences; Emanuelsson et al., 2003; Bodén and Hawkins, 2005; Hawkins et al., 2007). The quality of the generated data set was high, as validated by experimental analyses. Data set subgrouping further increased the quality of the data set used for model training (Figure 1A).

Data set–based discovery of so many plant PTS1 tripeptides was furthermore achieved by inclusion of several low-abundance proteins with atypical PTS1 tripeptides in the underlying set of known *Arabidopsis* PTS1 proteins. Most ESTs that were homologous to some low-abundance proteins, such as acetyl transferase 1/2 (ATF) or hydroxybutyryl-CoA dehydrogenase (HBCDH; Reumann et al., 2007) terminated with noncanonical and often novel PTS1 tripeptides. By contrast, the putative plant orthologs of high-abundance enzymes involved in photorespiration or fatty acid β-oxidation nearly all carry well-known canonical tripeptides and hardly contributed to the identification of novel PTS1s (Reumann, 2004; see Supplemental Data Set 1 online). Although the ESTs with noncanonical PTS1s presently remained low in relative and absolute numbers (see Supplemental Figure 1A online), they were highly instrumental in deducing novel functional plant PTS1 tripeptides (Figure 1).

Correct Inference of Novel PTS1 Tripeptides

Further PTS1 tripeptides were identified by our discriminative prediction models, omission of any PTS1 tripeptide filter, and by the models' ability to correctly infer novel PTS1 tripeptides. The recognition of noncanonical PTS1 tripeptides in low-abundance proteins identified by proteome analyses of plant peroxisomes (see Introduction) strongly suggested that the absence of a PTS1 tripeptide filter is an essential model property for predicting the entire proteome of plant peroxisomes. Both of our algorithms (PWM and RI models) combine the C-terminal PTS1 tripeptide and the upstream region (up to 12–amino acid residues) into a single prediction model. The models thereby exhibit a unique ability to correctly infer novel PTS1 tripeptides while maintaining high prediction specificity. The PWM model in particular is even able to correctly predict novel PTS1 tripeptide residues.

In terms of prediction sensitivity, the RI model presently seems to be too exclusive (i.e., insensitive). This can be explained by the higher model complexity of RI models, which allows them to represent and learn very subtle features of training sequences but also requires a larger training data set for best generalization performance (i.e., the ability to correctly predict unseen sequences) than the corresponding PWM models. Therefore, the simpler PWM model shows better generalization performance on this training data set of 2500 sequences. These observations call into question the accuracy of complex models that have been previously trained based on small data sets (90 to 300 sequences) for predicting novel PTS1 proteins (Emanuelsson et al., 2003; Bodén and Hawkins, 2005; Hawkins et al., 2007).

Although significantly superior in PTS1 protein prediction sensitivity on unseen sequences compared with the RI model, the PWM model should still be considered to be conservative. Five recently identified peroxisomal PTS1 proteins with noncanonical PTS1 tripeptides were scored below the threshold (see Supplemental Data Set 2 online). Additionally, four *Arabidopsis* proteins that we either demonstrated to possess functional PTS1 domains (CPK1, LKL> and PAP7, AHL>; Figure 5) or validated to be peroxisome targeted as full-length protein fusions in this study (NUDT19, SSL> and pxPfkB, SML>; Figure 5) were missed in the prediction of PTS1 proteins by this PWM model. Within an upper range of 1100 proteins in the hierarchical list of PWM model-predicted PTS1 proteins with a prediction score of at least 0.130 (GR1, TNL>, score = 0.162, hit number 1013; PAP7, score = 0.130, hit number 1118), further *Arabidopsis* PTS1 proteins must be expected to be found. Such a prediction gray zone below the threshold is still highly valuable for experimental biologists. Out of the large number of functionally as yet unknown *Arabidopsis* gene models, specific proteins with interesting annotation (i.e., domain conservation), such as those associated with auxin or JA metabolism, can be analyzed computationally for PTS1 conservation in putatively orthologous plant ESTs and experimentally for subcellular targeting in vivo in a relatively straightforward fashion.

Relaxation of the Plant PTS1 Motif

This study confirms 23 newly and six previously predicted PTS1 tripeptides to be true plant PTS1s by in vivo subcellular targeting analysis and increases the number of known plant PTS1s from 28 to 51. The newly experimentally verified PTS1 tripeptides add another 16 residues ([FVGTLKI][GETFPQCY]F>) to the 16 position-specific residues of the previously reported plant PTS1 motif ([SAPC][RKNMSLH][LMIVY]>; Figure 1B), leading to 11 (position −3), 15 (position −2), and six (position −1) allowed amino acid residues in plant PTS1 tripeptides. These results reveal a pronounced relaxation of the plant PTS1 motif that significantly extends and obviously contradicts the previous description as small (position −3), basic (position −2), and hydrophobic (position −1), particularly in positions −3 and −2. The basic position −2, which was previously considered to be the most conservative amino acid residue, is, based on our results, actually the most flexible, with 15 possible residues allowed out of 20 (75%), even including the acidic residue Glu (Figure 1B).

It is reasonable to predict that the number of plant PTS1 tripeptides and tripeptide residues will further increase in the near future. For instance, seven additional closely related tripeptides (e.g., SNI>, CRM>, and FRL>; Table 1) were found in a significant number (≥3) of positive example sequences and remain to be validated experimentally. Moreover, the era of experimental research on low-abundance peroxisomal matrix proteins and characterization of their atypical PTS1 tripeptides has begun only recently. EST database searches for putatively orthologous plant sequences using the *Arabidopsis* proteins identified in this study (see Supplemental Table 5 online) and others with noncanonical PTS1s, such as *Arabidopsis* glutathione reductase (TNL>; Kataya and Reumann, 2010) and NADK3

(SRY>; Waller et al., 2010), will certainly allow the recognition of further noncanonical PTS1 tripeptides.

In addition to the experimentally validated plant PTS1 tripeptides, the PWM model predicts 34 additional tripeptides as being functional in peroxisome targeting. Likewise, on top of the 32 experimentally validated plant PTS1 tripeptide residues (Figure 1B), the PWM model predicts that 10 additional residues might be allowed in plant PTS1 tripeptides ([HKQR][IAVW][QR]>; see Supplemental Data Set 2 online), leading to the prediction of 15 (position −3), 19 (position −2), and 8 (position −1) possible amino acid residues. Notably, all experimentally validated and PWM model-predicted plant PTS1 tripeptides follow a distinct pattern, in which at least two high-abundance residues of presumably strong targeting strength ([SA][KR][LMI]>; see Supplemental Figure 1B online) are combined with one low-abundance PTS1 residue to yield functional plant PTS1 tripeptides (x[KR][LMI]>, [SA]y[LMI]>, and [SA][KR]z>; Figure 1B).

High Prediction Specificity

Prediction models of high sensitivity often falsely predict a high number of proteins as organelle targeted. However, despite our models' ability to predict novel PTS1 tripeptide residues, they were not compromised for specificity, as documented by several parameters. First, the total number of 392 predicted *Arabidopsis* gene models out of 35,385 (1.1%) is relatively small. Second, only 51 (5%) of all possible amino acid residue combinations (11*15*6 = 990; Figure 1B) have now been established as functional PTS1s. Third, for the newly identified noncanonical and weak PTS1 tripeptides, only a very specific subset of *Arabidopsis* proteins is predicted to be peroxisome targeted (e.g., 1 out of 10 ALL> proteins). The prediction and experimental in vivo peroxisome targeting of proteins with noncanonical tripeptides depends on the presence of targeting-enhancing patterns in the upstream domain, as shown by the prediction analysis of all possible PTS1-nonapeptides (see Supplemental Figure 4 online) and by the analysis of the *Arabidopsis* genome (see Supplemental Table 5 online). Both prediction algorithms have learned specific targeting-enhancing patterns in the domain upstream of the PTS1 tripeptide and recognize these as essential elements for peroxisome targeting by weak PTS1 tripeptides. Cytosolic and peroxisome targeting of different sequences terminating with the same noncanonical PTS1 tripeptide (e.g., two VKL> sequences and three SPL> sequences; Figures 2 and 5) is an inherent rather than discrepant feature of noncanonical PTS1 tripeptides (see below).

Despite the large number of correctly predicted *Arabidopsis* PTS1 proteins, some false predictions must still be anticipated. Due to the disadvantageous C-terminal location of PTS1s in nascent polypeptides, some functional PTS1s might be overruled by N-terminal targeting signals or internal nuclear localization signals (Neuberger et al., 2004). Additionally, the PTS1 domain of a few proteins might be inaccessible to the cytosolic PTS1 receptor, Pex5p, in vivo due to conformational constraints (Neuberger et al., 2004; Ma and Reumann, 2008). Multiple subcellular targeting prediction analyses, combined with in vivo localization studies of N- and C-terminally and/or internally placed reporter proteins, are recommended to overcome these prevailing predictive limitations.

Prediction Validation by in Vivo Subcellular Targeting Analysis

Because of the large effort involved in experimental testing, comprehensive large-scale experimental validations of genome-wide organelle targeting predictions have not previously been reported. To validate the prediction accuracy of our models, we complemented the computational study by in vivo subcellular localization analyses of a total of more than 50 representative reporter protein constructs. The experimental verification rate was high. The detection of peroxisome targeting by weak PTS1s could be significantly improved by tissue incubation at low temperature, which reduced the rate of reporter protein and/or plasmid degradation and made possible subcellular targeting analysis after extended times of gene expression and protein import.

The identification of functional PTS1 tripeptides by this study required only qualitative peroxisome localization results. However, differential data on peroxisome targeting efficiencies yielded further insights into the biology of protein targeting to peroxisomes. The observed differential efficiencies of PTS1 decapeptides in directing EYFP to peroxisomes appears to be related to several parameters. First, the efficiency at which EYFP was targeted to peroxisomes by PTS1 decapeptides compared with full-length proteins might have been reduced because residues −11 to −14 might contain additional targeting enhancing residues (Figure 3). Second, EYFP fusions of different decapeptides carrying the same PTS1 tripeptides and full-length proteins generally differ in conformation and PEX5p accessibility of the C-terminal domain, all of which likely affects peroxisome targeting efficiency. Third, and to our mind most importantly, PTS1 domains carrying noncanonical PTS1 tripeptides generally appear to be of lower peroxisome targeting efficiency compared with canonical PTS1 domains. Most noncanonical PTS1 decapeptides of positive example sequences investigated experimentally in this study derived from low-abundance peroxisomal proteins, such as SOX, hydroxyacid oxidase 1 (HAOX1), and ATF1/2 (see Supplemental Table 1 online). By definition, low-abundance proteins are expressed at low rate in vivo. It appears that slowly produced proteins tolerate weak targeting signals because these are sufficient for quantitative protein targeting to peroxisomes. Consequently, these proteins have been lacking evolutionary pressure in evolving stronger, more efficient targeting signals. Under native conditions, the promoter strength of low-abundance peroxisomal proteins matches the expression level and leads to quantitative protein targeting to peroxisomes. In a heterologous expression system from a strong constitutive promoter, however, the expression rate of low-abundance peroxisomal proteins carrying weak PTS1 decapeptides exceeds the peroxisome import efficiency and results in residual cytosolic background fluorescence.

Regarding the positive example sequences of the reliable data set (represented by ≥3 sequences), all PTS1 tripeptides subjected to experimental analysis were validated as peroxisome targeted. Among the sequences of the uncertain data sets, three sequences with suspected PTS1 tripeptides remained cytosolic (RKL>, SEM>, and SGI>; Table 1, Figure 2), notably consistent with their PWM model predictions. These sequences derived

from ESTs, consistent with our initial hypothesis that single pass EST sequencing might have resulted in erroneous C-terminal tripeptides and/or targeting enhancing patterns. For instance, due to the high number of example sequences terminating with SKL> (654, 26%) and the close codon similarities between S (position −3, AG[UC]) and R (AG[AG]), single nucleotide errors in SKL> sequences might have led to the two erroneous RKL> sequences.

Significance of the Prediction Tools for Genome Screens

The prediction tools for PTS1 proteins are valuable for basic cell biology in the model plant species *Arabidopsis*. The multiple means of prediction information (e.g., PWM and RI model prediction scores and posterior probabilities and PTS1 tripeptide identifications) facilitate the selection of unknown *Arabidopsis* proteins of interesting annotation and straightforward in vivo validation of predicted peroxisome targeting. The methods make possible the long-awaited prediction of low-abundance and inducible peroxisomal matrix proteins, which are difficult to identify by experimental approaches. Several low-abundance proteins have already been identified in this study. Two homologs of CHY1, which is involved in branched amino acid catabolism (Zolman et al., 2001), a Cys protease, a PfkB homolog (pxPfkB), and SDRc are now established as peroxisomal proteins. The latter two proteins had been previously suggested to be peroxisome targeted based on proteome data (SDRc, Reumann et al., 2007; pxPfkB, Eubel et al., 2008). NUDT19 is a member of the nudix hydrolase family. NUDT7 and RP2p are peroxisomal in mammals and act as diphosphatases that cleave esterified or free CoASH into acyl- or 4′-phosphopantetheine and 3′,5′-ADP, thereby regulating peroxisomal CoA homeostasis (Gasmi and McLennan, 2001; Ofman et al., 2006; Reilly et al., 2008).

Our validation of functional PTS1 domains in nine additional *Arabidopsis* proteins (Figure 5) is likely to uncover further peroxisome-targeted PTS1 proteins. CPK1 was previously reported to be peroxisome targeted as a C-terminal reporter protein construct (CPK1-GFP) by a mechanism that depends on two potential N-terminal acylation sites (Dammann et al., 2003; Coca and San Segundo, 2010), rather than by the PTS1 pathway and LKL>. Several of the newly established *Arabidopsis* PTS1 proteins are inducible by abiotic stresses, as deduced from publicly available microarray data (data not shown; www.genevestigator.com; Zimmermann et al., 2005). These proteins may have important functions in plant adaptation to environmental stress. Moreover, many predicted PTS1 proteins have annotated functions related to pathogen defense and have been validated as peroxisome-targeted (A.R. Kataya, C. Mwaanga, and S. Reumann, unpublished data; see Supplemental Data Set 2 online). Functional studies, such as reverse genetics and protein–protein interaction analyses, will yield insights into the physiological functions of these proteins and into novel metabolic and regulatory networks of plant peroxisomes.

Because our prediction models require little computational time and memory, they can be easily applied to fully and partially sequenced plant genomes, including various crop plants and monocotyledons, such as rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*), which is an emerging model plant for biofuel production. Although these methods have been developed in sensu stricto for spermatophyta, the PTS1 protein prediction algorithms are also expected to be largely applicable to mosses (e.g., *Physcomitrella*). Future studies are needed to address whether plant PTS1s are conserved, for instance, in algae (e.g., *Chlamydomonas*) and whether these prediction tools are applicable to microalgae. The prediction of peroxisome functions in unicellular algae is expected to yield valuable insights into the evolution of peroxisome functions in higher plants.

Conclusions

The most important features of our PWM prediction model are summarized as follows: (1) the correct inference of many novel plant PTS1 tripeptides, (2) the correct prediction of a large number of unknown low-abundance *Arabidopsis* PTS1 proteins that could not have been uncovered by any other subcellular prediction tools currently available, and (3) the specific detection of these PTS1 proteins among many nonperoxisomal *Arabidopsis* proteins carrying the same tripeptide. Although the prediction algorithms outperform previously published methods, they still need to be improved further. The fact that the training data set is still underrepresented in low-abundance proteins presently limits the accuracy of our predictions. The unique ability of the PWM model to correctly predict low-abundance proteins with as yet undiscovered PTS1 tripeptides opens up strategic doors for systematically refining subcellular targeting prediction tools. By combining experimental and computational methodology in a targeted iterative approach, as was initiated in this study, low-abundance proteins that are predicted as peroxisome-targeted can be systematically validated experimentally. By subjection of these proteins to EST database searches for putatively orthologous sequences, the training data set can be progressively extended, allowing continuous improvement of the models' predictions and model refinement. Although it presently showed inferior prediction accuracy on unknown proteins, the RI model is expected to reveal its full prediction potential on extended data sets generated by the proposed iterative strategy.

METHODS

Data Set Generation and the Discriminative Machine Learning Approach

The methodology is described in detail in the Supplemental Methods online.

In Vivo Subcellular Localization Studies

For validation of the data set and of the PTS1 domains that were predicted by the model, the C-terminal 10 residues of plant full-length cDNAs or ESTs (see Supplemental Table 1 online) were fused to the C terminus of EYFP by PCR using an extended reverse primer (see Supplemental Tables 1 and 7 online) and subcloned into the plant expression vector pCAT (Fulda et al., 2002) under control of a double 35S cauliflower mosaic virus promoter. To study the subcellular targeting of *Arabidopsis thaliana* full-length cDNAs with predicted PTS1s in plant cells, fusion proteins with N-terminally located EYFP were generated. *Arabidopsis* cDNAs were

ordered from the ABRC and the RIKEN Biosource Centre with primers containing appropriate restriction endonuclease sites (see Supplemental Table 6 online) and subcloned, in frame, into the same plant expression vector. All constructs were fully sequenced; single amino acid point mutations located distantly to the PTS1 domain were observed in CHY1H1 (At2g30650, 378 amino acids, K331R), CUT1 (At1g68530, 497 amino acids, I131T), and Cys protease (At3g57810, 317 amino acids, E199K and F297S). The sequences of all constructs are made available online as Fasta files (see Supplemental Data Sets 3 to 7 online). For labeling of peroxisomes in double transformants, a fusion protein of the N-terminal 50 residues of glyoxysomal malate dehydrogenase (CsgMDH) from *Cucumis sativus* comprising the PTS2 targeting domain and ECFP was used (CsgMDH-ECFP; Fulda et al., 2002). Onion epidermal cells were transformed biolistically as described (Ma et al., 2006). The onion slices were placed on wet paper in Petri dishes, stored at room temperature in the dark for ~16 h, and analyzed directly or after tissue incubation at 10°C for 1 to 6 d.

Image Capture and Analysis

Fluorescence image acquisition was performed on a Nikon TE-2000U inverted fluorescence microscope equipped with an Exfo X-cite 120 fluorescence illumination system and either single filters for YFP (exciter HQ500/20, emitter S535/30) and CFP (exciter D436/20, emitter D480/40) or a dual YFP/CFP filter with single-band exciters (Chroma Technologies). All images were captured using a Hamamatsu Orca ER 1394 cooled CCD camera. Standard image acquisition and analysis were performed using Volocity II software (Improvision) and Photoshop.

Accession Numbers

Accession numbers from this article can be found in Supplemental Table 5 online.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Statistical Analyses of Positive Example Sequences.

Supplemental Figure 2. Comparative Subcellular Targeting Results after Different Expression Times.

Supplemental Figure 3. Dependency of Posterior Probabilities on the Prediction Scores of the PWM and RI Models.

Supplemental Figure 4. PWM and RI Model-Based Predictions of Peroxisome Targeting for PTS1 Tripeptides with all Possible Combinations of Upstream Hexapeptides.

Supplemental Figure 5. Distribution of *Arabidopsis* Gene Models and Loci by Their Prediction Scores as Peroxisome-Targeted PTS1 Proteins.

Supplemental Table 1. Sequence Information, Prediction Data, and Experimental Validation Results of Positive Example Sequences.

Supplemental Table 2. Performance Comparison of Two Discriminative Prediction Models for Plant PTS1 Proteins.

Supplemental Table 3. PWM Performance Regarding Alternative Residue Order and Sequence Redundancy Reduction.

Supplemental Table 4. Most Discriminative Features of the PTS1 Protein Prediction Models.

Supplemental Table 5. Protein Information, Prediction Data, and Experimental Validation Results of Representative *Arabidopsis* Proteins.

Supplemental Table 6. Oligonucleotide Primers Used for cDNA Subcloning.

Supplemental Table 7. List of Acronyms of PTS1 Proteins and Plant Species Investigated Experimentally.

Supplemental Methods.

Supplemental Data Set 1. PTS1 Protein Prediction Scores for Positive and Negative Example Sequences.

Supplemental Data Set 2. PWM and RI Model-Based PTS1 Protein Predictions for 35,386 *Arabidopsis* Gene Models (TAIR 10).

Supplemental Data Sets 3 to 7. Fasta Files.

ACKNOWLEDGMENTS

We thank the *Arabidopsis* stock centers ABRC and RIKEN for the provision of full-length cDNAs and Nora Valeur for subcloning help. We also thank Jianping Hu for critical reading of the manuscript. S.R. and T.L. were supported by fellowships from Lower Saxony and the DAAD Post-Doc programme, respectively. The research was supported by the Deutsche Forschungsgemeinschaft and the University of Stavanger.

Received February 4, 2011; revised February 4, 2011; accepted March 24, 2011; published April 12, 2011.

REFERENCES

Arai, Y., Hayashi, M., and Nishimura, M. (2008a). Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. Plant Cell Physiol. **49:** 526–539.

Arai, Y., Hayashi, M., and Nishimura, M. (2008b). Proteomic identification and characterization of a novel peroxisomal adenine nucleotide transporter supplying ATP for fatty acid beta-oxidation in soybean and *Arabidopsis*. Plant Cell **20:** 3227–3240.

Babujee, L., Wurtz, V., Ma, C., Lueder, F., Soni, P., van Dorsselaer, A., and Reumann, S. (2010). The proteome map of spinach leaf peroxisomes indicates partial compartmentalization of phylloquinone (vitamin K1) biosynthesis in plant peroxisomes. J. Exp. Bot. **61:** 1441–1453.

Bodén, M., and Hawkins, J. (2005). Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics **21:** 2279–2286.

Bongcam, V., MacDonald-Comber Petétot, J., Mittendorf, V., Robertson, E.J., Leech, R.M., Qin, Y.M., Hiltunen, J.K., and Poirier, Y. (2000). Importance of sequences adjacent to the terminal tripeptide in the import of a peroxisomal *Candida tropicalis* protein in plant peroxisomes. Planta **211:** 150–157.

Brocard, C., and Hartig, A. (2006). Peroxisome targeting signal 1: Is it really a simple tripeptide? Biochim. Biophys. Acta **1763:** 1565–1573.

Coca, M., and San Segundo, B. (2010). AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. Plant J. **63:** 526–540.

Dammann, C., Ichida, A., Hong, B., Romanowsky, S.M., Hrabak, E.M., Harmon, A.C., Pickard, B.G., and Harper, J.F. (2003). Subcellular targeting of nine calcium-dependent protein kinase isoforms from Arabidopsis. Plant Physiol. **132:** 1840–1848.

Distel, B., Gould, S.J., Voorn-Brouwer, T., van der Berg, M., Tabak, H.F., and Subramani, S. (1992). The carboxyl-terminal tripeptide serine-lysine-leucine of firefly luciferase is necessary but not sufficient for peroxisomal import in yeast. New Biol. **4:** 157–165.

Emanuelsson, O., Elofsson, A., von Heijne, G., and Cristóbal, S. (2003). In silico prediction of the peroxisomal proteome in fungi, plants and animals. J. Mol. Biol. **330:** 443–456.

Eubel, H., Meyer, E.H., Taylor, N.L., Bussell, J.D., O'Toole, N., Heazlewood, J.L., Castleden, I., Small, I.D., Smith, S.M., and Millar, A.H. (2008). Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of Arabidopsis cell culture peroxisomes. Plant Physiol. **148:** 1809–1829.

Fukao, Y., Hayashi, M., Hara-Nishimura, I., and Nishimura, M. (2003). Novel glyoxysomal protein kinase, GPK1, identified by proteomic analysis of glyoxysomes in etiolated cotyledons of *Arabidopsis thaliana*. Plant Cell Physiol. **44:** 1002–1012.

Fukao, Y., Hayashi, M., and Nishimura, M. (2002). Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana*. Plant Cell Physiol. **43:** 689–696.

Fulda, M., Shockey, J., Werber, M., Wolter, F.P., and Heinz, E. (2002). Two long-chain acyl-CoA synthetases from *Arabidopsis thaliana* involved in peroxisomal fatty acid beta-oxidation. Plant J. **32:** 93–103.

Gasmi, L., and McLennan, A.G. (2001). The mouse Nudt7 gene encodes a peroxisomal nudix hydrolase specific for coenzyme A and its derivatives. Biochem. J. **357:** 33–38.

Goepfert, S., Hiltunen, J.K., and Poirier, Y. (2006). Identification and functional characterization of a monofunctional peroxisomal enoyl-CoA hydratase 2 that participates in the degradation of even cis-unsaturated fatty acids in *Arabidopsis thaliana*. J. Biol. Chem. **281:** 35894–35903.

Hawkins, J., Mahony, D., Maetschke, S., Wakabayashi, M., Teasdale, R.D., and Bodén, M. (2007). Identifying novel peroxisomal proteins. Proteins **69:** 606–616.

Hayashi, M., and Nishimura, M. (2003). Entering a new era of research on plant peroxisomes. Curr. Opin. Plant Biol. **6:** 577–582.

Kataya, A.R., and Reumann, S. (2010). Arabidopsis glutathione reductase 1 is dually targeted to peroxisomes and the cytosol. Plant Signal. Behav. **5:** 171–175.

Kaur, N., Reumann, S., and Hu, J. (2009). Peroxisome Biogenesis and Function. In The Arabidopsis Book **7:** e0123, doi/10.1199/tab.0123.

Kragler, F., Lametschwandtner, G., Christmann, J., Hartig, A., and Harada, J.J. (1998). Identification and analysis of the plant peroxisomal targeting signal 1 receptor NtPEX5. Proc. Natl. Acad. Sci. USA **95:** 13336–13341.

Lipka, V., et al. (2005). Pre- and postinvasion defenses both contribute to nonhost resistance in *Arabidopsis*. Science **310:** 1180–1183.

Lisenbee, C.S., Lingard, M.J., and Trelease, R.N. (2005). Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. Plant J. **43:** 900–914.

Lopez-Huertas, E., Charlton, W.L., Johnson, B., Graham, I.A., and Baker, A. (2000). Stress induces peroxisome biogenesis genes. EMBO J. **19:** 6770–6777.

Ma, C., Haslbeck, M., Babujee, L., Jahn, O., and Reumann, S. (2006). Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. Plant Physiol. **141:** 47–60.

Ma, C., and Reumann, S. (2008). Improved prediction of peroxisomal PTS1 proteins from genome sequences based on experimental subcellular targeting analyses as exemplified for protein kinases from *Arabidopsis*. J. Exp. Bot. **59:** 3767–3779.

Mintz-Oron, S., Aharoni, A., Ruppin, E., and Shlomi, T. (2009). Network-based prediction of metabolic enzymes' subcellular localization. Bioinformatics **25:** i247–i252.

Mitschke, J., Fuss, J., Blum, T., Höglund, A., Reski, R., Kohlbacher, O., and Rensing, S.A. (2009). Prediction of dual protein targeting to plant organelles. New Phytol. **183:** 224–235.

Moschou, P.N., Sanmartin, M., Andriopoulou, A.H., Rojo, E., Sanchez-Serrano, J.J., and Roubelakis-Angelakis, K.A. (2008). Bridging the gap between plant and mammalian polyamine catabolism: A novel peroxisomal polyamine oxidase responsible for a full back-conversion pathway in Arabidopsis. Plant Physiol. **147:** 1845–1857.

Mullen, R.T., Lee, M.S., Flynn, C.R., and Trelease, R.N. (1997). Diverse amino acid residues function within the type 1 peroxisomal targeting signal. Implications for the role of accessory residues upstream of the type 1 peroxisomal targeting signal. Plant Physiol. **115:** 881–889.

Nair, R., and Rost, B. (2008). Protein subcellular localization prediction using artificial intelligence technology. Methods Mol. Biol. **484:** 435–463.

Neuberger, G., Kunze, M., Eisenhaber, F., Berger, J., Hartig, A., and Brocard, C. (2004). Hidden localization motifs: Naturally occurring peroxisomal targeting signals in non-peroxisomal proteins. Genome Biol. **5:** R97.

Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003a). Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. J. Mol. Biol. **328:** 567–579.

Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003b). Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J. Mol. Biol. **328:** 581–592.

Nyathi, Y., and Baker, A. (2006). Plant peroxisomes as a source of signalling molecules. Biochim. Biophys. Acta **1763:** 1478–1495.

Ofman, R., Speijer, D., Leen, R., and Wanders, R.J. (2006). Proteomic analysis of mouse kidney peroxisomes: Identification of RP2p as a peroxisomal nudix hydrolase with acyl-CoA diphosphatase activity. Biochem. J. **393:** 537–543.

Pain, D., Schnell, D.J., Murakami, H., and Blobel, G. (1991). Machinery for protein import into chloroplasts and mitochondria. Genet. Eng. (N. Y.) **13:** 153–166.

Picard, R., and Cook, D. (1984). Cross-validation of regression models. J. Am. Stat. Assoc. **79:** 575–583.

Purdue, P.E., and Lazarow, P.B. (2001). Peroxisome biogenesis. Annu. Rev. Cell Dev. Biol. **17:** 701–752.

Quan, S., Switzenberg, R., Reumann, S., and Hu, J. (2010). In vivo subcellular targeting analysis validates a novel peroxisome targeting signal type 2 and the peroxisomal localization of two proteins with putative functions in defense in Arabidopsis. Plant Signal. Behav. **5:** 151–153.

Reilly, S.J., Tillander, V., Ofman, R., Alexson, S.E., and Hunt, M.C. (2008). The nudix hydrolase 7 is an Acyl-CoA diphosphatase involved in regulating peroxisomal coenzyme A homeostasis. J. Biochem. **144:** 655–663.

Reumann, S. (2004). Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol. **135:** 783–800.

Reumann, S. (2011). Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics. Proteomics **11:** 1764–1779.

Reumann, S., Babujee, L., Ma, C., Wienkoop, S., Siemsen, T., Antonicelli, G.E., Rasche, N., Lüder, F., Weckwerth, W., and Jahn, O. (2007). Proteome analysis of *Arabidopsis* leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell **19:** 3170–3193.

Reumann, S., Ma, C., Lemke, S., and Babujee, L. (2004). AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. Plant Physiol. **136:** 2587–2608.

Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C.G., Weber,

A.P., Olsen, L.J., and Hu, J. (2009). In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol. **150:** 125–143.

Reumann, S., and Weber, A.P. (2006). Plant peroxisomes respire in the light: Some gaps of the photorespiratory C2 cycle have become filled—others remain. Biochim. Biophys. Acta **1763:** 1496–1510.

Rifkin, R., Yeo, G., and Poggio, T. (2003). Regularized Least Squares Classification In Advances in Learning Theory: Methods, Model and Applications. NATO Science Series III: Computer and Systems Sciences, J.A.K. Suykens, I. Horvath, S. Basu, C. Micchelli, and J. Vandewalle, eds (Amsterdam: IOS Press), pp. 131–153.

Schlüter, A., Real-Chicharro, A., Gabaldón, T., Sánchez-Jiménez, F., and Pujol, A. (2010). PeroxisomeDB 2.0: An integrative view of the global peroxisomal metabolome. Nucleic Acids Res. **38** (Database issue): D800–D805.

Schneider, G., and Fechner, U. (2004). Advances in the prediction of protein targeting signals. Proteomics **4:** 1571–1580.

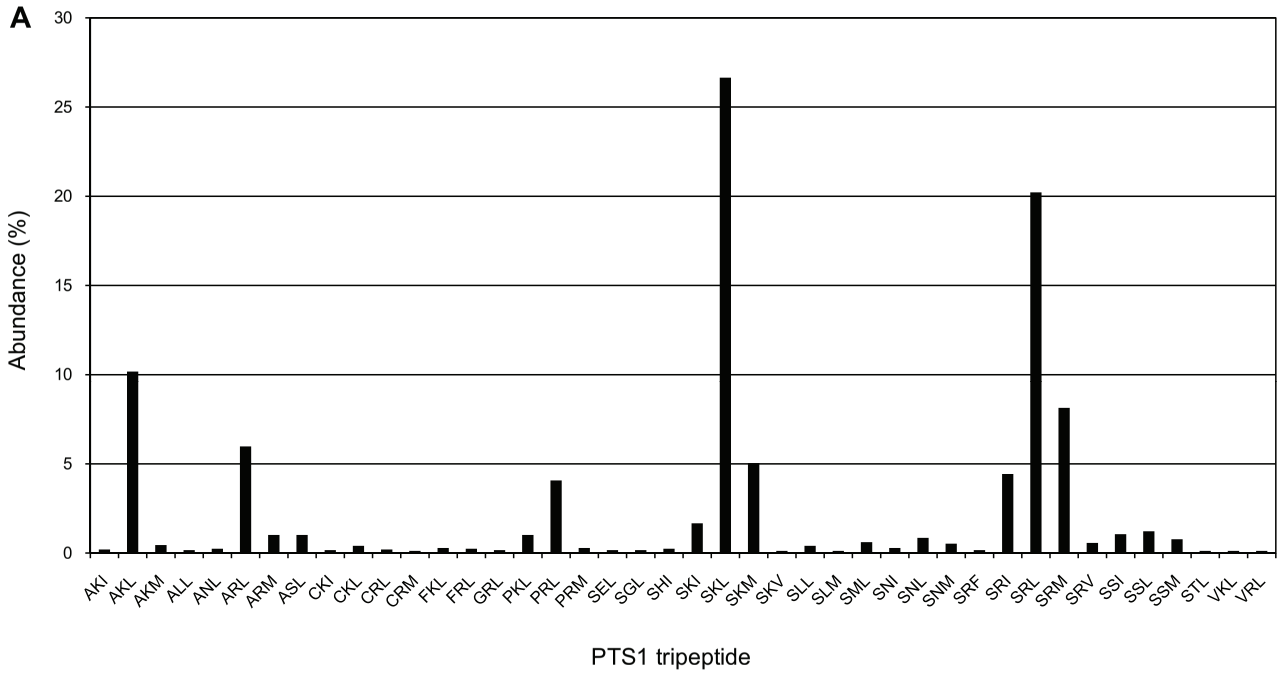
Schnell, D.J., and Hebert, D.N. (2003). Protein translocons: Multifunctional mediators of protein translocation across membranes. Cell **112:** 491–505.

Waller, J.C., Dhanoa, P.K., Schumann, U., Mullen, R.T., and Snedden, W.A. (2010). Subcellular and tissue localization of NAD kinases from *Arabidopsis*: Compartmentalization of de novo NADP biosynthesis. Planta **231:** 305–317.

Zimmermann, P., Hennig, L., and Gruissem, W. (2005). Gene-expression analysis and network discovery using Genevestigator. Trends Plant Sci. **10:** 407–409.

Zolman, B.K., Monroe-Augustus, M., Thompson, B., Hawes, J.W., Krukenberg, K.A., Matsuda, S.P., and Bartel, B. (2001). chy1, an *Arabidopsis* mutant with impaired beta-oxidation, is defective in a peroxisomal beta-hydroxyisobutyryl-CoA hydrolase. J. Biol. Chem. **276:** 31037–31046.

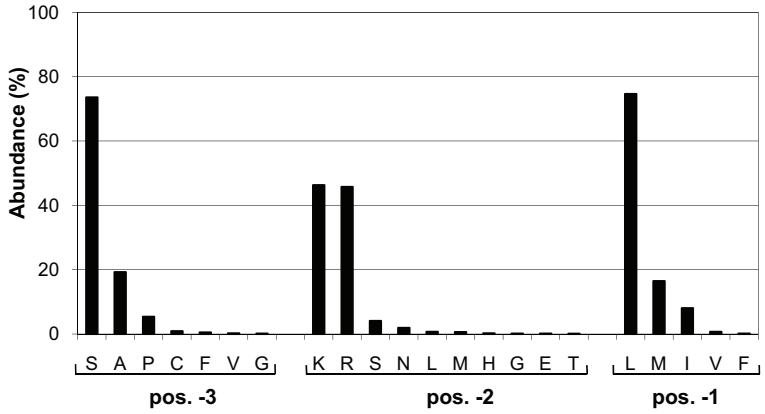
Suppl. Fig. 1



Suppl. Fig. 1. Statistical analyses of positive example sequences. The positive example sequences of data subset 1 were analyzed for abundance of the 42 PTS1 tripeptides (A).

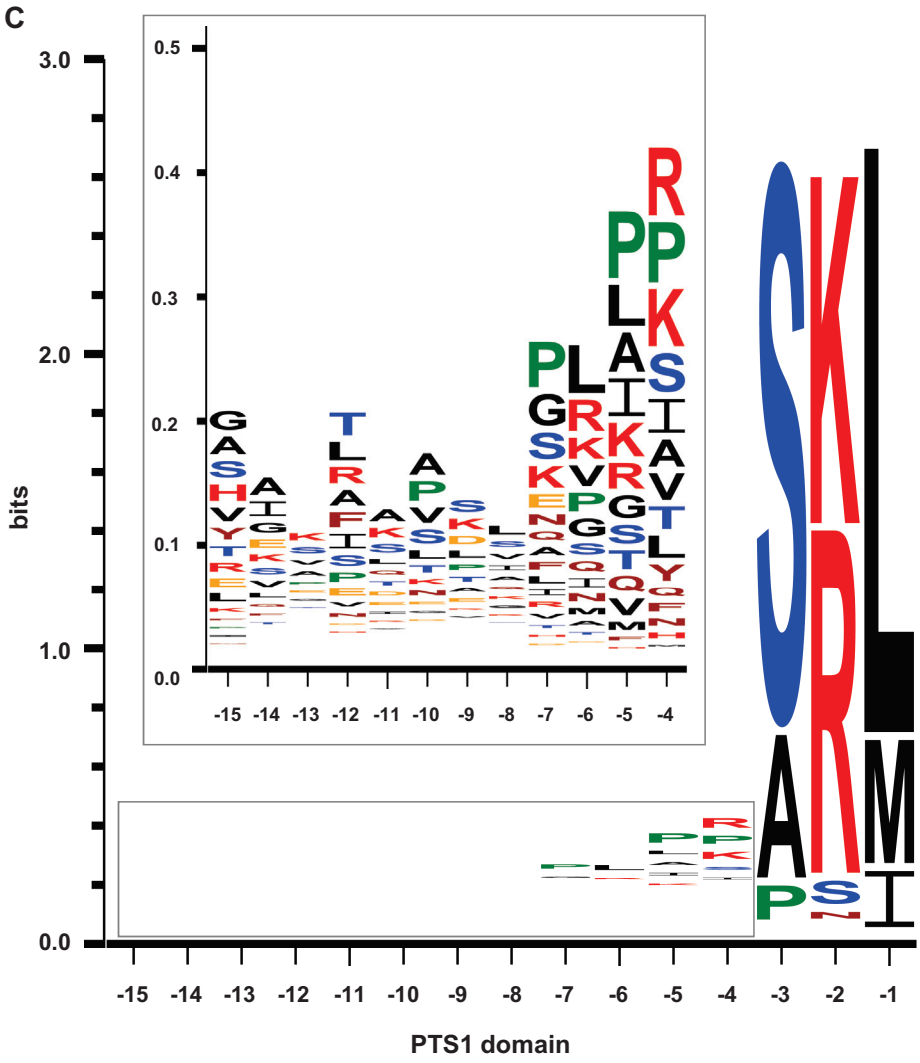
Suppl. Fig. 1

B



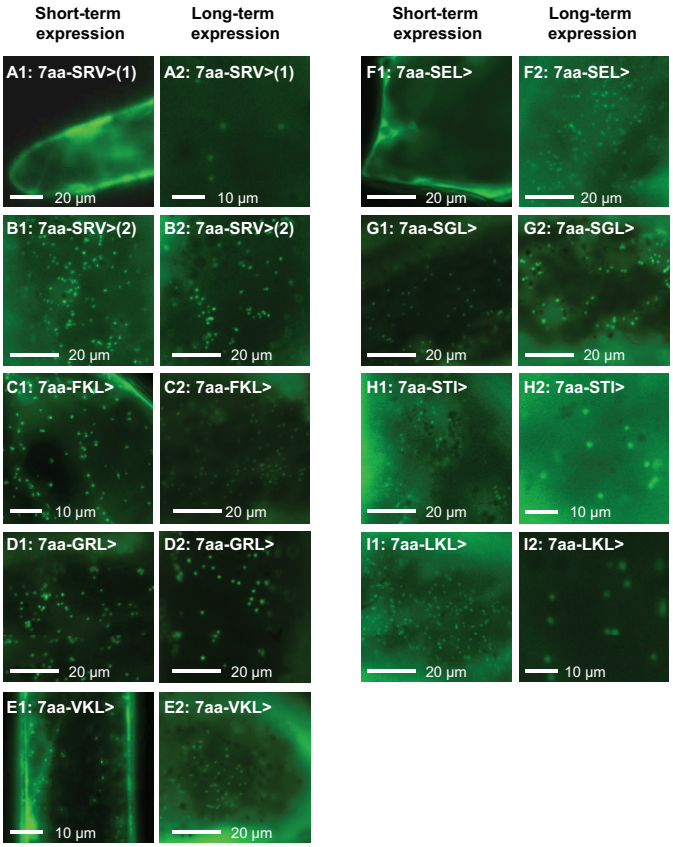
Suppl. Fig. 1. Statistical analyses of positive example sequences. The positive example sequences of data subset 1 were analyzed for abundance of the position-specific residues in the tripeptides (B).

Suppl. Fig. 1



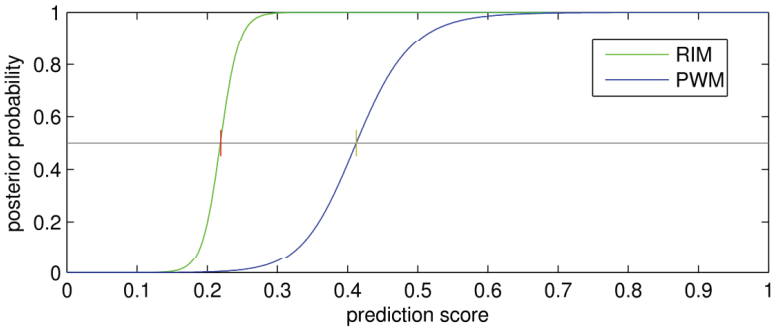
Suppl. Fig. 1. Statistical analyses of positive example sequences. The positive example sequences of data subset 1 were analyzed for abundance of aa residues in the upstream domain using web logo (C, http://weblogo.berkeley.edu/logo.cgi). The following color code was used for the aa residues: K, R, and H (red); D and E (orange); S and T (blue); P (green); G, A, L, M, I, and V (black); Q, N, F, Y, W, and C (brown).

Suppl. Fig. 2



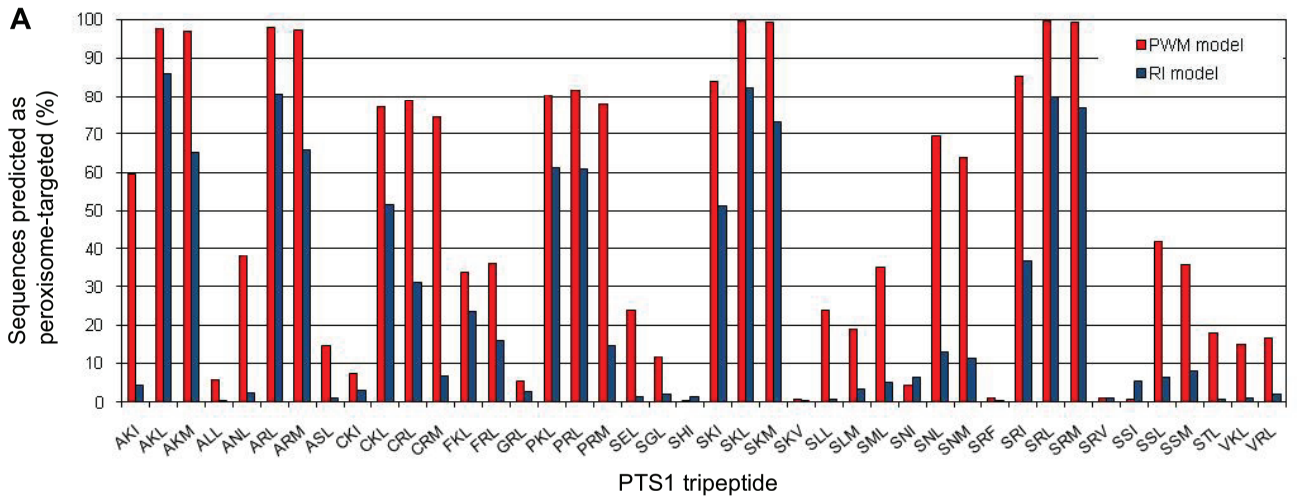
Suppl. Figure 2 online. Comparative subcellular targeting results after different expression times. Onion epidermal cells were transformed biolistically with EYFP fusion constructs that were C-terminally extended by the C-terminal decapeptides of plant PTS1 proteins serving as example sequences. Subcellular targeting was analyzed comparatively by fluorescence microscopy after short-term expression, i.e. either ca. 18 h expression at RT only (B1-E1, G1) or an additional 24 h at ca. 10°C (A1, F1, H1, I1), or after long-term expression, i.e., an additional five days of expression at ca. 10°C (A2-I2). To document the efficiency of peroxisome targeting, EYFP images of single transformants were not modified for brightness or contrast. One of each image set is also shown in Fig. 2. For sequence details, see Suppl. Tables 1 and 7 online.

Suppl. Fig. 3



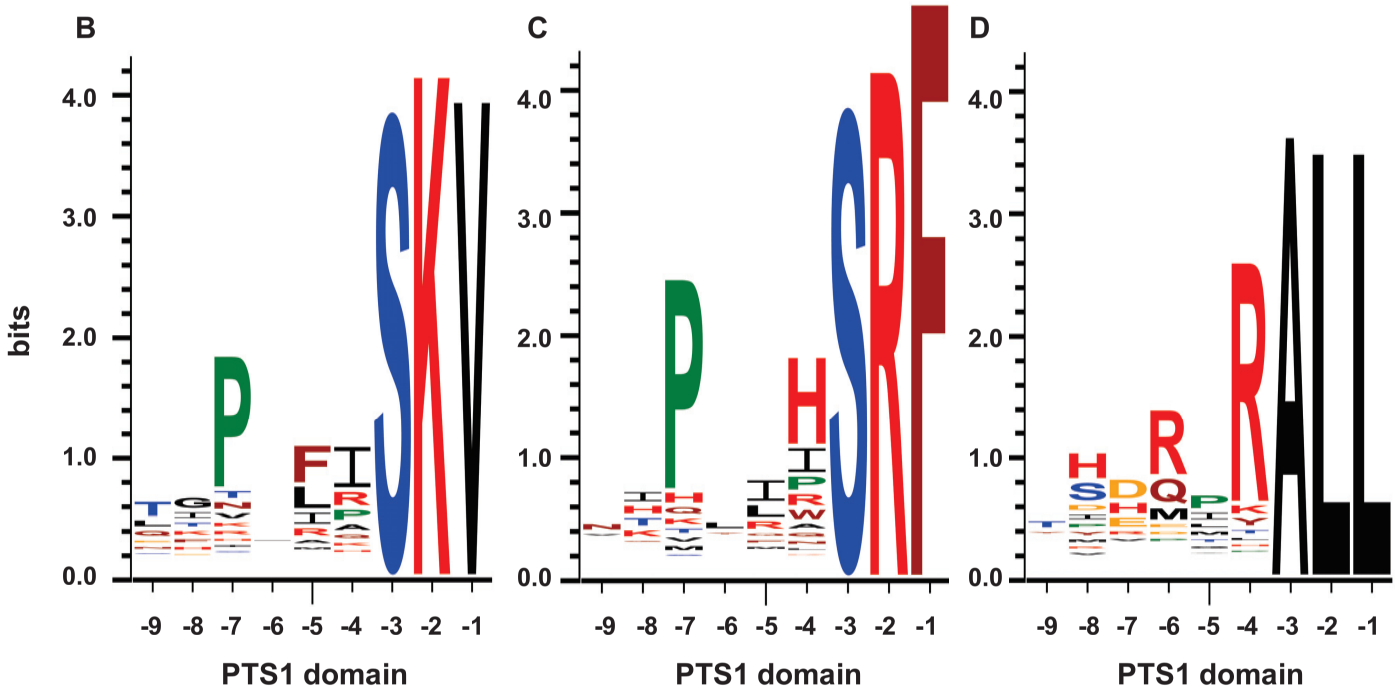
Suppl. Fig. 3. Dependency of posterior probabilities on the prediction scores of the PWM and RI models. The red and yellow lines indicate the prediction scores that correspond to a posterior probability of 0.5 for the RI and PWM model, respectively.

Suppl. Fig. 4



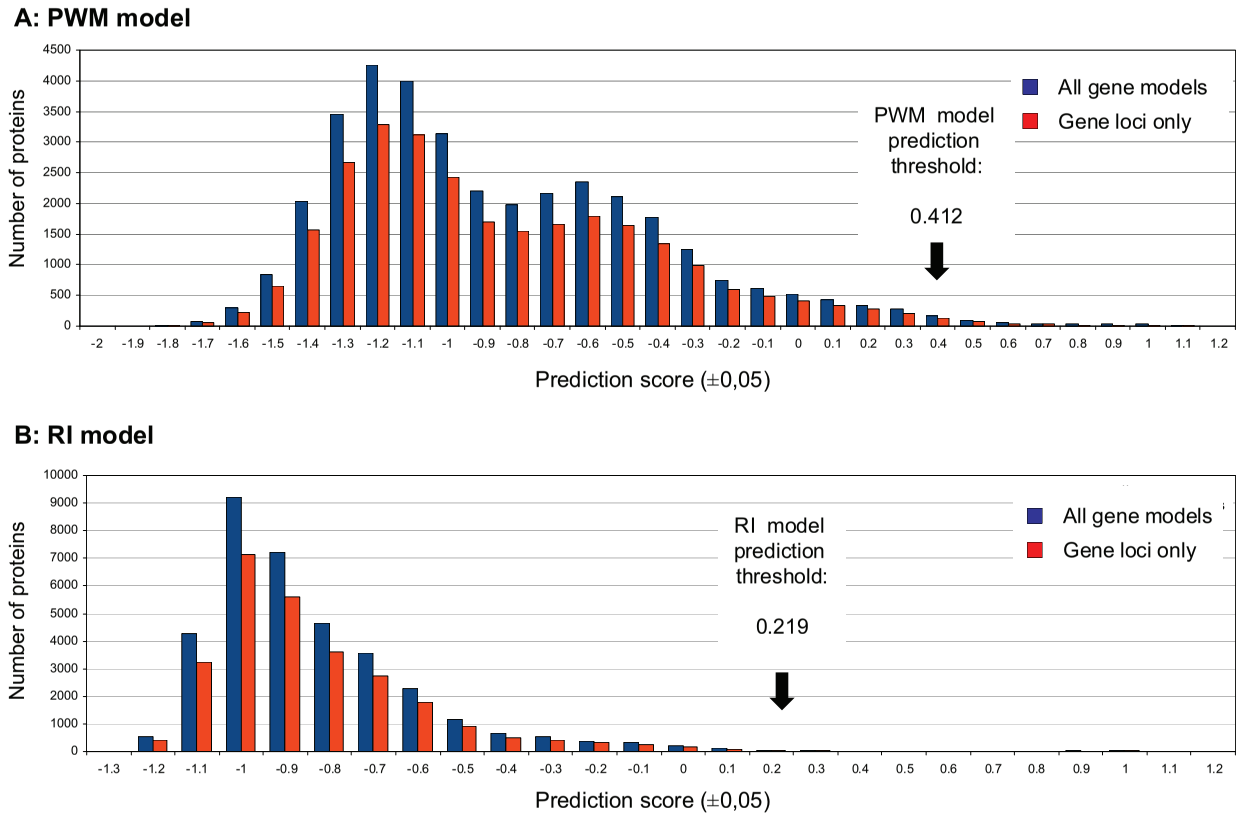
Suppl. Fig. 4. PWM and RI model-based predictions of peroxisome targeting for PTS1 tripeptides with all possible combinations of upstream hexapeptides. The relative numbers of upstream hexapeptides predicted as peroxisome-targeted are presented (A).

Suppl. Fig. 4



Suppl. Fig. 4. PWM and RI model-based predictions of peroxisome targeting for PTS1 tripeptides with all possible combinations of upstream hexapeptides. Web logo presentations are shown for nonapeptides that are predicted by the RI model as peroxisome-targeted and terminate with SKV> (B), SRF> (C), or ALL> (D). For the three PTS1 tripeptides, 105,610 (SKV>), 161,875 (SRF>), and 88,252 nonapeptides (ALL>) are predicted to be peroxisome-targeted. The following color code was used for the aa residues: K, R, and H (red); D and E (orange); S and T (blue); P (green); G, A, L, M, I, and V (black); Q, N, F, Y, W, and C (brown).

Supp. Fig. 5



Suppl. Fig. 5. Distribution of *Arabidopsis* gene models and loci by their prediction scores as peroxisome-targeted PTS1 proteins.
The distribution is shown for the PWM (A) and the RI prediction model (B).

Suppl. Table 1: Sequence information, prediction data, and experimental validation results of positive example sequences.

The sequences are sorted by the prediction scores and posterior probabilities of the PWM model. For plant protein and species acronyms see Suppl. Table 7. TS: training sequences, US: unseen sequence(s), fl: full-length cDNA, P: peroxisome, C: cytosol

PTS1 protein	Spe-cies	Sequence type	C-terminal decapeptide fused to EYFP	Sequence number (TS/US)	PWM prediction model			RI prediction model			Experimental subcellular targeting	Fig. 2
					Prediction score	Post. probability	Targeting prediction	Prediction score	Post. probability	Targeting prediction		
SOX	Lse	EST	VHIQVRH**SSM>**	19 (TS)	0.886	0.999	P	0.857	1.0	P	Peroxisome	D
HAOX1	At	fl	TENERIK**SML>**	15 (TS)	0.782	0.999	P	0.969	1.0	P	Peroxisome	B
HAOX1	Os	EST	TEGDRIR**SLL>**	10 (TS)	0.776	0.999	P	1.045	1.0	P	Peroxisome	L
SOX	Gh	EST	VQVRVGH**SNM>**	13 (TS)	0.775	0.999	P	1.012	1.0	P	Peroxisome	C
AAE7	Gg	EST	KMGPVRK**CKL>**	10 (TS)	0.698	0.998	P	0.592	1.0	P	Peroxisome	G
SOX	St	fl	VHVRVGQ**ANL>**	6 (TS)	0.652	0.995	P	0.847	1.0	P	Peroxisome	F
GOX	Phv	EST	RTHPRTR**ALL>**	4 (TS)	0.641	0.994	P	0.504	1.0	P	Peroxisome	N
MFP2	Cro	EST	SQINQAK**SQL>**	2 (US)	0.577	0.978	P	0.092	0.0	C	Peroxisome	X
GOX	Cro	EST	DTSPRTK**STL>**	3 (TS)	0.572	0.976	P	0.569	1.0	P	Peroxisome	T
AGT	Ptri xPd	fl	NNTPLIA**SRV>(2)**	14 (TS)	0.505	0.902	P	0.813	1.0	P	Peroxisome	I
AGT	Pt	EST	KTTPMIP**SRV>(3)**	14 (TS)	0.482	0.845	P	0.772	1.0	P	Peroxisome	J
ECoAI	Gr	EST	QKRKPTF**VKL>**	3 (TS)	0.458	0.754	P	0.602	1.0	P	Peroxisome	Q
Uri	Ze	EST	ASLSRTL**FKL>**	7 (TS)	0.440	0.665	P	0.597	1.0	P	Peroxisome	M
AIM1	Vu	EST	SSTPKLL**GRL>**	4 (TS)	0.435	0.636	P	0.495	1.0	P	Peroxisome	P
ATF1	Nt	EST	LHKEDLK**SHI>**	6 (TS)	0.418	0.535	P	0.912	1.0	P	Peroxisome	K
DEG15	Rs	EST	LSRDVIP**SEL>**	4 (TS)	0.410	0.485	C	0.375	1.0	P	Peroxisome	R
SOX	So	EST	LATPDLR**SFM>**	2 (US)	0.396	0.398	C	0.165	0.011	C	Peroxisome	Z
AtPAO2	Aa	EST	ASIPLLI**SRF>**	4 (TS)	0.375	0.276	C	0.579	1.0	P	Peroxisome	V
SCADHb	Gr	EST	TPVGVPS**RKL>**	2 (US)	0.367	0.233	C	0.139	0.001	C	Cytosol	Ae
GOX	Zm	EST	DKFSAIP**SGL>**	4 (TS)	0.359	0.200	C	0.426	1.0	P	Peroxisome	S
sT5	Sme	EST	SLFNKLR**SKV>**	3 (TS)	0.318	0.074	C	0.613	1.0	P	Peroxisome	E
MDAR1	Vv	fl	KEGLTFA**CKI>**	4 (TS)	0.281	0.027	C	0.755	1.0	P	Peroxisome	O
SCP	Fv	EST	SDIFPKP**SEM>**	2 (US)	0.270	0.020	C	0.252	0.936	P	Cytosol	Ad
ECHIc	Vv	fl	GEVSHTR**TRL>**	2 (US)	0.269	0.019	C	0.012	0.0	C	Peroxisome	Ab
PAO3	Ccl	EST	ISVPFLI**SPL>**	2 (US)	0.247	0.010	C	0.056	0.0	C	Peroxisome	W

ACX4	Ze	EST	VAKTTRP**SRV>(1)**	14 (TS)	0.129	0.000	C	0.379	1.0	P	Peroxisome	H
AGT	Gr	EST	NNIPMSP**SGI>**	1 (US)	0.175	0.001	C	-0.095	0.0	C	Cytosol	Ac
ATF	Psi	fl	YLTLKLL**PKI>**	2 (US)	0.135	0.0	C	-0.189	0.0	C	Peroxisome	Aa
GSTT1	AfxAp	EST	SSIKAML**STI>**	1 (US)	0.055	0.0	C	-0.269	0.0	C	Peroxisome	U
ECoAI	Hv	EST	RSRGPFS**LKL>**	1 (US)	0.021	0.0	C	-0.038	0.0	C	Peroxisome	Y
Uri	So	EST	TSLDPPM**LNL>**	2 (US)	-0.161	0.0	C	-0.501	0.0	C	Cytosol	Af
GOX_b	Pn	EST	TCSRWDH**LCR>**	2 (US)	-1.433	0.0	C	-0.967	0.0	C	Cytosol	Ag

Suppl. Table 2: Performance comparison of two discriminative prediction models for plant PTS1 proteins. The performance values are based on the C-terminal 14 (PWM model) and 15 (RI model) aa residues and were derived from 3-fold cross validation. Sensitivity and specificity values correspond to the fraction of true positive example sequences, i.e., a sensitivity of 0.98 corresponds to 98% correctly detected positive examples.

Performance index	PWM model	RI model
Sensitivity	0.981	0.996
Specificity	0.959	0.970
Harmonic mean	0.970	0.983
Evaluation time for 21028	0.34 s	0.37 s
sequences	≤ 0.016 ms/seq.	≤ 0.018 ms/seq.
Score range of pos. examples		
maximum	1.376	1.347
minimum	0.032	0.375
Score range of neg. examples		
maximum	1.182	0.766
minimum	-1.855	-1.264
Prediction threshold	0.412	0.219

Suppl. Table 3: PWM performance regarding alternative residue order and sequence redundancy reduction. Upstream residues were added in reverse order to subsequently growing models (A). Average and standard deviation of 50 repetitions of random residue sampling are presented, whereby six upstream residues were randomly selected from the upstream decapeptide (B). Datasets were reduced to 90% sequence similarity (C). For further method details see also Suppl. Methods.

A) reversed order of upstream residues

length of C-terminus	sensitivity	specificity	harmonic mean
4	0.967	0.916	0.941
5	0.969	0.922	0.945
6	0.971	0.927	0.949
7	0.975	0.954	0.964
8	0.976	0.959	0.967
9	0.971	0.963	0.967
10	0.972	0.965	0.968
11	0.973	0.965	0.969
12	0.974	0.966	0.970
13	0.971	0.967	0.969
14	0.973	0.968	0.970
15	0.975	0.969	0.972

B) randomly selected upstream residues (50 repetitions)

	sensitivity	specificity	harmonic mean
average	0.972	0.955	0.964
standard deviation	0.002	0.012	0.006

C) datasets reduced to 90% sequence similarity

length of C-terminus	sensitivity	specificity	harmonic mean
3	0.958	0.917	0.937
4	0.962	0.921	0.941
5	0.964	0.926	0.945
6	0.965	0.929	0.947
7	0.966	0.931	0.948
8	0.965	0.935	0.950
9	0.965	0.937	0.951
10	0.963	0.940	0.952
11	0.964	0.942	0.953
12	0.965	0.942	0.953
13	0.966	0.943	0.954
14	0.966	0.945	0.956
15	0.965	0.946	0.956

Suppl. Table 4: Most discriminative features of the PTS1 protein prediction models. The 20 most discriminative features of the PWM (A) and RI model (B) are presented according to values of the discriminative weight vector associated with the best-performing model. Column 1 denotes the rank, and columns 2-4 and 5-7 correspond to the associated positive and negative features, respectively. Features are represented as aa (in IUPAC notation) at particular positions (PWM model, A) and aa pairs at particular positions (RI model, B), respectively. The corresponding discriminative weight is given.

A: PWM model

#	Positive features			Negative features		
	Aa	Position	Weight	Aa	Position	Weight
1	L	-1	0.66	W	-6	-0.33
2	M	-1	0.64	Y	-11	-0.28
3	S	-3	0.48	W	-3	-0.26
4	R	-2	0.46	T	-1	-0.24
5	K	-2	0.44	R	-1	-0.23
6	A	-3	0.34	D	-3	-0.23
7	I	-1	0.33	C	-1	-0.22
8	W	-14	0.15	A	-1	-0.21
9	W	-13	0.15	K	-1	-0.21
10	P	-3	0.13	W	-1	-0.21
11	C	-3	0.12	V	-2	-0.20
12	M	-6	0.07	G	-3	-0.20
13	R	-6	0.03	W	-8	-0.19
14	R	-4	0.03	C	-9	-0.19
15	P	-5	0.03	S	-1	-0.19
16	P	-10	0.03	F	-2	-0.19
17	H	-4	0.03	I	-3	-0.19
18	D	-11	0.02	L	-3	-0.19
19	P	-7	0.02	P	-1	-0.19
20	Y	-9	0.02	H	-1	-0.18

B: RI model

#	Positive features					Negative features				
	Upstream domain		PTS1 tripeptide		Weight	Upstream domain		PTS1 tripeptide		Weight
	Aa	Pos.	Aa	Pos.		Aa	Pos.	Aa	Pos.	
1	P	-5	L	-1	0.06	D	-4	S	-3	-0.03
2	K	-8	S	-3	0.06	L	-12	L	-1	-0.03
3	K	-7	L	-1	0.06	D	-4	L	-1	-0.03
4	P	-7	R	-2	0.05	F	-7	L	-1	-0.03
5	P	-7	S	-3	0.05	Y	-11	S	-3	-0.02
6	K	-4	M	-1	0.05	Y	-11	L	-1	-0.02
7	I	-5	S	-3	0.05	D	-5	S	-3	-0.02
8	L	-5	S	-3	0.05	K	-15	A	-3	-0.02
9	K	-4	S	-3	0.05	E	-4	L	-1	-0.02
10	R	-4	A	-3	0.05	G	-10	S	-2	-0.02
11	A	-11	L	-1	0.05	G	-15	S	-2	-0.02
12	R	-4	R	-2	0.05	S	-7	S	-2	-0.02
13	L	-5	I	-1	0.05	L	-5	E	-2	-0.02
14	S	-4	L	-1	0.04	I	-11	S	-3	-0.02
15	K	-4	I	-1	0.04	Y	-5	S	-3	-0.02
16	R	-4	L	-1	0.04	F	-4	S	-3	-0.02
17	E	-14	S	-3	0.04	D	-6	L	-1	-0.02
18	R	-4	M	-1	0.04	D	-6	R	-2	-0.02
19	W	-14	S	-3	0.04	F	-4	I	-1	-0.02
20	R	-4	S	-3	0.04	W	-6	S	-3	-0.02

Suppl. Table 5: Protein information, prediction data, and experimental validation results of representative *Arabidopsis* proteins.

The *Arabidopsis* proteins investigated experimentally are grouped by the type of experimental analysis (EYFP-PTS1 domain constructs or EYFP full-length cDNA constructs) and sorted by the PWM-based prediction scores and posterior probabilities. Subcellular targeting predictions to peroxisomes (P) or the cytosol (C) are indicated. [1, 2] ACS3 (At5g28360.1) and the C-terminal domain of UP9 (At2g48060.1) investigated experimentally had been predicted by TAIR9 and TAIR8, respectively, while no protein or a different C-terminal end are predicted by TAIR10. [3]SRY> and Tyr (Y, pos. -1) had in parallel been demonstrated to be a novel PTS1 tripeptide and an allowed PTS1 residue (NADK3, Waller et al., 2010). Novel PTS1 residues (in addition to those identified in Fig. 2) are underlined. The ability of the prediction models to correctly infer novel PTS1 tripeptides is indicated. Regarding the prediction accuracy of pre-existing tools, only four proteins (SDRc, PHD, PAP7, and NUDT19) out of eleven newly identified *Arabidopsis* proteins carrying non-canonical PTS1 tripeptides (plus UP9, Tudor, SPK1, LCAT, ACS3, CPK1, pxPfkB) had been correctly predicted as peroxisomal by the PTS1 predictor, and none of the ten proteins had been correctly predicted by PeroxiP. n.a., not applicable (because the PTS1 tripeptide was included in the training dataset subset 1 or the fusion protein was cytosolic, Fig. 1).

AGI code	Acronym	Annotation	C-terminal 10 aa residues	PWM prediction model				RI prediction model				Experimental subcellular targeting	Fig. 5
				Pred. score	Post. prob.	Targ. pred.	PTS1 infer-ence	Pred. score	Post. prob.	Targ. pred.	PTS1 infer-ence		
EYFP-PTS1 domain constructs													
At3g20530.1	PK1	Protein kinase 1	EEEEDER**SKL**>	0.890	1.0	P	n.a.	0.461	1.0	P	n.a.	peroxisome	P
At2g48060.1	UP9	Unknown protein[2]	TYGTTLP**S<u>C</u>L**>[2]	0.722	0.998	P	yes	0.281	0.993	P	yes	peroxisome	A
At1g02190.1/2	FAH	Fatty acid hydrolase superfamily	DFQPLPP**SPL**>(1)	0,707	0,998	P	n.a.	0,081	0.0	C	n.a.	cytosol	B
At2g40640.1	RING	RING/U-box superfamily protein	TVAEMPP**SWL**>	0,627	0,992	P	n.a.	0,158	0,006	C	n.a.	cytosol	D
At1g51745.1/2	Tudor	Tudor/PWWP/MBT superfamily protein	RQQRRQR**<u>K</u>RL**>	0,615	0,990	P	yes	0,160	0,007	C	no	peroxisome	E
At3g01980.1/3/4	SDRc	Short-chain dehydro-genase/reductase c	LTRPRLK**<u>S</u>YM**>	0,610	0,989	P	yes	0,344	1,000	P	yes	peroxisome	F
At4g16340.1	SPK1	Guanyl-nucleotide exchange factor	HYIPAIL**SEL**>	0,567	0,973	P	n.a.	0,176	0,027	C	n.a.	peroxisome	I, L
At1g43770.2	PHD	RING/FYVE/PHD zinc finger superfamily protein	VFKPRQT**SR<u>Y</u>**>[3]	0,499	0,891	P	yes	0,110	0,000	C	no	peroxisome	H
At2g14255.1	ANK	Ankyrin repeat family protein with DHHC zinc finger domain	DDEVVPS**SIL**>	0,446	0,700	P	n.a.	0,162	0,008	C	n.a.	cytosol	J
At3g44830.1	LCAT	Lecithin:cholesterol acyltransferase family protein	RMSERIS**<u>I</u>KL**>	0,438	0,657	P	yes	0,200	0,177	C	no	peroxisome	K

AGI code	Acronym	Annotation	C-terminal 10 aa residues	PWM prediction model				RI prediction model				Experimental subcellular targeting	Fig. 5
				Pred. score	Post. prob.	Targ. pred.	PTS1 infer-ence	Pred. score	Post. prob.	Targ. pred.	PTS1 infer-ence		
At5g28360.1	ACS3[1]	1-Aminocyclopropane-1-carboxylate synthase like pseudogene	ENYLRLV**SPL**>(2)	0.426	0.582	P	n.a.	-0.058	0.0	C	n.a.	peroxisome	C
At5g04870.1	CPK1	Ca-dep. protein kinase 1	EKSFSIA**LKL**>	0,321	0,080	C	no	-0,165	0.0	C	no	peroxisome	N
At1g68530.1	CUT1	Cuticular 1 / 3-Ketoacyl-CoA synthase 6	PVFIPEV**VKL**>	0,293	0,037	C	n.a.	-0,190	0.0	C	n.a.	cytosol	O
At2g01880.1	PAP7	Purple acid phosphatase 7	SSLSKRS**AHL**>	0,130	0.0	C	no	-0,162	0.0	C	no	peroxisome	M
At3g01980.2	SDRc	Short-chain dehydro-genase/reductase c	QLVRVCL**APN**>	-0,767	0.0	C	n.a.	-0,828	0.0	C	n.a.	cytosol	G
EYFP full-length protein constructs													
At3g57810.1/2/3	CP	Cysteine proteinases superfamily protein	SKASIPK**SKL**>	0,998	1.0	P	n.a.	0,739	1.0	P	n.a.	peroxisome	Q
At2g30660.1	CHY1H2	CHY1 homolog 2 / ATP-dependent caseinolytic (Clp) protease/crotonase family protein	LHALRIA**AKL**>	0,924	1.0	P	n.a.	0,640	1.0	P	n.a.	peroxisome	S
At2g30650.1	CHY1H1	CHY1 homolog 1 / ATP-dependent caseinolytic (Clp) protease/crotonase family protein	LPALAIA**AKL**>	0,682	0,997	P	n.a.	0,384	1.0	P	n.a.	peroxisome	R
At3g01980.1/3/4	SDRc	Short-chain dehydro-genase/reductase c	LTRPRLK**SYM**>	0,610	0,989	P	yes	0,344	1.0	P	yes	peroxisome	T
At4g36195.1/2	S28FP	Serine carboxypeptidase S28 family protein	ECRGGIR**SSM**>	0,534	0,946	P	n.a.	0,078	0,0	C	n.a.	non-peroxisomal structures	U
At5g20070.1	NUDT19	Nudix hydrolase homolog 19	QQAGVSL**SSL**>	0,385	0,328	C	n.a.	0,054	0,0	C	n.a.	peroxisome	V
At1g49350.1	pxPfkB	(peroxisomal) PfkB-type carbohydrate kinase family protein	KMLMVHQ**SML**>	0,298	0,044	C	n.a.	0,114	0,0	C	n.a.	peroxisome	W
At1g68530.1	CUT1	Cuticular 1 / 3-Ketoacyl-CoA synthase 6	PVFIPEV**VKL**>	0,293	0,037	C	n.a.	-0,190	0,0	C	n.a.	cytosol	X

Suppl. Table 6: Oligonucleotide primers used for cDNA subcloning

One forward primer was used for EYFP amplification and introduced a 5'-NcoI site into the PCR products (5'-AAGTCCATGGTGAGCAAGGGCGAGGA-3'). The primers are sorted alphabetically according to the construct name. [1, 2] ACS3 (At5g28360.1) and the C-terminal domain of UP9 (At2g48060.1) investigated experimentally had been predicted by TAIR9 and TAIR8, respectively, while no protein or a different C-terminal end are predicted by TAIR10. Protein and plant species acronyms are explained in Suppl. Table 7.

Construct	PTS1 protein	AGI code	Plant species	Nucleotide source	C-terminal 10 aa residues	Primer	Primer sequence (5' to 3')
PTS1 domain primers of example sequences							
EYFP-7aa-ALL>	GOX		Phv	EST	RTHPRTRALL>	rev	TATATCTAGATCACAATAACGCACGGGTGCGAGGGTGAGTGCGCTTGTACAGCTCGTCCATGCC
EYFP-7aa-ANL>	SOX		St	flcDNA	VHVRVGQANL>	rev	TATATCTAGACTAAAGATTTGCTTGACCAACTCGGACATGAACCTTGTACAGCTCGTCCATGCC
EYFP-7aa-CKI>	MDAR1		Vv	genomic/ flcDNA	KEGLTFACKI>	rev	TATATCTAGAGTCAAATCTTACAGGCAAAGGTGAGACCTTCCTTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-CKL>	AAE7		Gg	EST	KMGPVRKCKL>	rev	TATATCTAGAGTCATAGTTTGCATTTCCGGACGGGCCCCATCTTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-FKL>	Uri		Ze	EST	ASLSRTLFKL>	rev	TATATCTAGACTACAATTTGAATAGGGTGCGGCTCAAGCTAGCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-GRL>	AIM1		Vu	EST	SSTPKLLGRL>	rev	TATATCTAGAGTCATAGACGCCCCAGTAGCTTTGGAGTAGATGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-LCR>	GOX_b		Pn	EST	TCSRWDHLCR>	rev	TATATCTAGAGTCAGCGACACAGATGGTCCCATCGGCTACATGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-LKL>	ECoAI		Hv	EST	RSRGPFSLKL>	rev	TATATCTAGAGTCAGAGCTTTAGAGAAAACGGGCCTCTGCTTCTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-LNL>	Uri		So	EST	TSLDPPMLNL>	rev	TATATCTAGAGTCACAGATTTAACATGGGCGGGTCAAGGGATGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-PKI>	ATF		Psi	genomic/ flcDNA	YLTLKLLPKI>	rev	TATATCTAGAGTCATATCTTTGGTAGTAATTTCAACGTCAAGTACTTGTACAGCTCGTCCATGCC
EYFP-7aa-RKL>	SCADHb		Gr	EST	TPVGVPSRKL>	rev	TATATCTAGAGTCACAGCTTGCGGGAGGGAACCCCCACCGGGGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SEL>	DEG15		Rs	EST	LSRDVIPSEL>	rev	TATATCTAGAGTCATAACTCGCTAGGGATCACGTCGCGGGAAAGCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SEM>	SCP		Fv	EST	SDIFPKPSEM>	rev	TATATCTAGAGTCACATCTCGGAAGGCTTGGGGAAAATGTCGGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-SFM>	SOX		So	EST	LATPDLRSFM>	rev	TATATCTAGATCACATGAAGCTTCTAAGATCAGGTGTAGCCAACTTGTACAGCTCGTCCATGCC
EYFP-7aa-SGI>	AGT		Gr	EST	NNIPMSPSGI>	rev	TATATCTAGAGTCAAATCCCTGAAGGGGACATTGGGATGTTATTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SGL>	GOX		Zm	EST	DKFSAIPSGL>	rev	TATATCTAGAGTCACAAGCCCGACGGGATGGCGCTGAACTTGTCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SHI>	ATF1		Nt	EST	LHKEDLKSHI>	rev	TATATCTAGAGTCAAATGTGAGATTTCAGGTCCTCTTTGTGCAACTTGTACAGCTCGTCCATGCC
EYFP-7aa-SKV>	sT5		Sme	EST	SLFNKLRSKV>	rev	TATATCTAGAGTCAAACTTTGCTTCTTAGTTTATTAAATAACGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-SLL>	HAOX1		Os	EST	TEGDRIRSLL>	rev	TATATCTAGATCAGAGCAGGGACCTGATCCTGTCACCCTCGGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SML>	HAOX1	At3g14130.1	At	genomic/ flcDNA	TENERIKSML>	rev	TATATCTAGATCAGAGCATAGATTTAATTCTCTCATTCTCTGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SNM>	SOX		Gh	EST	VQVRVGHSNM>	rev	TATATCTAGATTACATATTTGAGTGCCCCACGCGAACTTGAACCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SPL>	PAO3		Ccl	EST	ISVPFLISPL>	rev	TATATCTAGAGTCACAGAGGGGAGATCAAAAATGGGACAGAAATCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SQL>	MFP2		Cro	EST	SQINQAKSQL>	rev	TATATCTAGAGTCACAGCTGGGACTTGGCCTGGTTTATCTGAGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-SRF>	AtPAO2		Aa	EST	ASIPLLISRF>	rev	TATATCTAGAGTCAAAAACGAGATATCAGTAAAGGAATAGAAGCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SRV>(1)	ACX4		Ze	EST	VAKTTRPSRV>	rev	TATATCTAGAGTCACACGCGGCTGGGGCGGGTCGTCTTTGCAACCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SRV>(2)	AGT		PtrixPd	genomic/ flcDNA	NNTPLIASRV>	rev	TATGTCTAGAGTCAaaccctggaagcgatcagaggggtattattCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SRV>(3)	AGT		Pt	EST	KTTPMIPSRV>	rev	TATGTCTAGAGTCAcaccctagatggtatcatgggagtggttttCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SSM>	SOX		Lse	EST	VHIQVRHSSM>	rev	TATATCTAGATTACATGCTTGAATGCCTAACTTGTATATGAACCTTGTACAGCTCGTCCATGCC
EYFP-7aa-STI>	GSTT1		AfxAp	EST	SSIKAMLSTI>	rev	TATATCTAGAGTCATATTGTTGACAACATGGCCTTTATGCTTGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-STL>	GOX		Cro	EST	DTSPRTKSTL>	rev	TATATCTAGAGTCATAACGTGGACTTAGTACGAGGAGAAGTATCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-TRL>	ECHIc		Vv	genomic/ flcDNA	GEVSHTRTRL>	rev	TATATCTAGAGTCATAGCCGTGTGCGAGTGTGACTCACTTCTCCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-VKL>	ECoAI		Gr	EST	QKRKPTFVKL>	rev	TATATCTAGATTAAAGCTTAACAAAAGTTGGTTTCCTTTTTTGCTTGTACAGCTCGTCCATGCC

Construct	PTS1 protein	AGI code	Plant species	Nucleotide source	C-terminal 10 aa residues	Primer	Primer sequence (5' to 3')
PTS1 domain primers of newly predicted Arabidopsis proteins							
EYFP-7aa-AHL>	PAP7	At2g01880.1	At	cDNA	SSLSKRSAHL>	rev	TATGTCTAGAGTCAAAGATGAGCAGATCTTTTGGACAATGAGGACTTGTACAGCTCGTCCATGCC
EYFP-7aa-APN>	SDRc	At3g01980.2	At	cDNA	LVRVCILAPN>	rev	TATGTCTAGAGTCAATTAGGAGCTAGGATGCATACCCGGACGAGCTTGTACAGCTCGTCCATGCC
EYFP-7aa-IKL>	LCAT	At3g44830.1	At	cDNA	RMSERISIKL>	rev	TATGTCTAGAGTCACAACTTGATGCTAATTCTCTCCGACATCCTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-KRL>	Tudor	At1g51745.1	At	cDNA	RQQRRQRKRL>	rev	TATGTCTAGAGTCACAACCGTTTCCTTTGCCTTCTCTGTTGTCTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-LKL>	CPK1	At5g04870.1	At	cDNA	EKSFSIALKL>	rev	TATGTCTAGAGTCAGAGTTTAAGAGCAATGCTAAAGCTTTTCTCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SCL>	UP9[1]	At2g48060.1[1]	At	cDNA	TYGTTLPSCL>	rev	TATGTCTAGAGTCAAAGACAAGAAGGAAGTGTTGTACCGTAAGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SEL>	SPK1	At4g16340.1	At	cDNA	HYIPAILSEL>	rev	TATGTCTAGAGTCAAAGTTCGGATAAGATAGCAGGGATGTAATGCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SIL>	ANK	At2g14255.1	At	cDNA	DDEVVPSSIL>	rev	TATGTCTAGAGTCAGAGAATTGAAGACGGGACAACCTCGTCGTCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SPL>(1)	FAH	At1g02190.1	At	cDNA	DFQPLPPSPL>	rev	TATGTCTAGAGTCATAGAGGAGATGGTGGGAGAGGTTGGAAATCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SPL>(2)	ACS3[2]	At5g28360.1	At	cDNA	ENYLRLVSPL>	rev	TATGTCTAGAGTCACAAAGGGCTCACAAGTCTCAAGTAATTCTCCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SRY>	PHD	At1g43770.2	At	cDNA	VFKPRQTSRY>	rev	TATGTCTAGAGTCAATATCGAGAAGTCTGTCTAGGCTTGAAGACCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SYM>	SDRc	At3g01980.1/ 3/4	At	cDNA	LTRPRLKSYM>	rev	TATGTCTAGAGTCACATGTAGGATTTGAGACGGGGTCGCGTAAGCTTGTACAGCTCGTCCATGCC
EYFP-7aa-SWL>	RING	At2g40640.1	At	cDNA	TVAEMPPSWL>	rev	TATGTCTAGAGTCATTAGAGCCAAGACGGTGGCATTTCAGCTACTGTCTTGTACAGCTCGTCCATGCC
EYFP-7aa-VKL>	CUT1	At1g68530.1	At	cDNA	PVFIPEVVKL>	rev	TATGTCTAGAGTCAGAGTTTGACAACTTCGGGAATAAAGACAGGCTTGTACAGCTCGTCCATGCC
Primers of newly predicted Arabidopsis full-length cDNAs							
EYFP-CHY1H1	CHY1H1	At2g30650.1	At	cDNA	LPALAIAAKL>	for	AAGACTGCGGCCGCTATGGCATCTCACTCTCAGGTT
						rev	CAAGTCTAGAGTCAAAGCTTTGCTGCGATTGC
EYFP-CHY1H2	CHY1H2	At2g30660.1	At	cDNA	LHALRIAAKL>	for	AAGACTGCGGCCGCTATGGCATCTCAATCTCAGGTT
						rev	CAAGCCGCGGTCACAACTTTGCTGCAATTCT
EYFP-CP	CP	At3g57810.1	At	cDNA	SKASIPKSKL>	for	AAGACTGCGGCCGCTATGATGATTTGTTACTCTCCA
						rev	CAAGTCTAGAGCTAAAGTTTAGATTTTGGAATCGA
EYFP-CUT1	CUT1	At1g68530.1	At	cDNA	PVFIPEVVKL>	for	AAGACTGCGGCCGCTATGCCTCAGGCACCGATGCCA
						rev	CAAGTCTAGAGTTAGAGTTTGACAACTTCGGG
EYFP-NUDT19	NUDT19	At5g20070.1	At	cDNA	QQAGVSLSSL>	for	AAGACTGCGGCCGCTATGCTTGCTCTCTTCCTCTCT
						rev	CAAGCCGCGGCTACAAACTTGAGAGAGAGAC
EYFP-pxPfkB	pxPfkB	At1g49350.1	At	cDNA	KMLMVHQSML>	for	AAGACTGCGGCCGCTATGGAACCGGTTATAATCGGT
						rev	CAAGTCTAGAGTCATAGCATCGACTGGTGAAC
EYFP-S28FP	S28FP	At4g36195.1	At	cDNA	ECRGGIRSSM>	for	AAGACTGCGGCCGCTATGTTGTCGGCGCTTGGCTTT
						rev	CAAGCCGCGGTTACATGGAACTCCTAATCCC
EYFP-SDRc	SDRc	At3g01980.1	At	cDNA	LTRPRLKSYM>	for	AAGACTGCGGCCGCTATGGAAAATCCGGCGAAGAGA
						rev	CAAGCCGCGGTCACATGTAGGATTTGAGACG

Suppl. Table 7: List of acronyms of PTS1 proteins and plant species investigated experimentally.

A. Protein acronyms

PTS1 protein	**Name**
AAE7	Acyl activating enzyme isoform 7
ACS3	1-Aminocyclopropane-1-carboxylate synthase like pseudogene
ACX4	Acyl-CoA oxidase isoform 4
AGT	Alanine-glyoxylate aminotransferase (SGT)
AIM1	ABNORMAL INFORESCENCE 1
ANK	Ankyrin repeat family protein with DHHC zinc finger domain
ATF / ATF1	Acetyl transferase 1
CER1	CER1 protein
CHY1H1	CHY1 homolog 1
CHY1H2	CHY1 homolog 2
CP	Cysteine protease family protein
CPK1	Ca-dep. protein kinase 1
CUT1	Cuticular 1
DEG15	DEG15 protease
ECHIc	Enoyl-CoA hydratase/isomerase isoform c
ECoAI	Enoyl-CoA hydratase/isomerase isoform
FAH	Fatty acid hydrolase superfamily
GOX	Glycolate oxidase
GSTT1	Glutatione-S transferase isoform theta
HAOX1	Hydroxyacid oxidase 1
LCAT	Lecithin:cholesterol acyltransferase family protein
MDAR1	Monodehydroascorbate reductase isoform 1
MFP2	Multifunctional protein
NUDT19	Nudix hydrolase homolog 19
PAO2	Polyamine oxidase 2
PAO3	Polyamine oxidase isoform 3
PAP7	Purple acid phosphatase 7
pxPfkB	(peroxisomal) PfkB-type carbohydrate kinase family protein
PHD	PHD finger family protein
RING	RING/U-box superfamily protein
S28FP	Serine carboxypeptidase S28 family protein
SCADHb	Short-chain dehydrogenase/reductase b
SCP	Sterol carrier protein
SDRc	Short-chain dehydrogenase/reductase c
SOX	Sulfite oxidase
SPK1	GTP binding protein
sT5	Small thioesterase isoform 5
Tudor	Tudor/PWWP/MBT superfamily protein
UP9	Unknown protein 9
Uri	Uricase
ZnBP	Zinc ion binding protein

B. Plant species acronyms

Acronym	**Plant species**
Aa	*Acorus americanus*
AfxAp	*Aquilegia formosa* x *aquilegia pubescens*
At	*Arabidopsis thaliana*
Ccl	*Citrus clementina*
Cro	*Catharanthus roseus* (madagascar periwinkle)
Fv	*Fragaria vesca*
Gg	*Gnetum gnemon*
Gh	*Gossypium hirsutum* (upland cotton)
Gr	*Gossypium raimondii*
Hv	*Hordeum vulgare*
Lse	*Lactuca serriola*
Nt	*Nicotiana tabacum* (common tobacco)
Os	*Oryza sativa*
Phv	*Phaseolus vulgaris*
Pn	*Populus nigra*
Psi	*Picea sitchensis* (sitka spruce)
Pt	*Pinus taeda* (loblolly pine)
Ptri x Pd	*Populus trichocarpa* x *Populus deltoides*
Rs	*Raphanus sativus* (radish)
Sme	*Saussurea medusa*
So	*Saccharum officinarum*
St	*Solanum tuberosum* (potato)
Vu	*Vigna unguiculata* (cowpea)
Vv	*Vitis vinifera*
Ze	*Zinnia elegans*
Zm	*Zea mays*

Supplemental Methods

Dataset generation. To generate a large and representative dataset of positive example sequences for PTS1 proteins of spermatophyta, we collected 2,562 peroxisomal proteins from known proteins and ESTs. Sixty known Arabidopsis PTS1 proteins were used as queries against the nonredundant database to investigate whether putative orthologs from other plant species could be identified unambiguously based on maximum sequence similarity and minimum Expect value by uni- and bidirectional BLAST searches and putatively orthologous protein sequences retrieved. For suitable PTS1 proteins, a minimum level of sequence similarity in the C-terminal 100 aa residues was defined in order to discriminate between PTS1 protein orthologs and paralogs. The same threshold was used to search plant EST databases for putatively orthologous sequences. For ambiguous sequences, putative orthology was further confirmed by bi-directional BLAST and multiple sequence alignment analyses, and uncertain sequences were disregarded. Protein and EST sequences were retrieved irrespective of the identity of their C-terminal tripeptides. According to their C-terminal tripeptides, the example sequences were grouped into three datasets (see Fig. 1A) and are made available online as fasta files (Supplemental Datasets 5-7).

To generate a sufficiently large and representative set of negative example sequences (i.e., non-peroxisomal proteins) for the discriminative learning approach, all protein sequences from spermatophyta (Taxonomy ID 58024) within the manually curated SwissProt database (release 56.5, November 2008) were initially collected. From this set, sequences that were annotated with SwissProt keywords “Peroxisome” (KW 576) or “Glyoxysome” (KW 330) and sequences that are associated with Gene Ontology terms “peroxisome” (GO 0005777) or “glyoxysome” (GO 0009514) were excluded. Further sequences of less than 20 amino acids, or sequences that contained undetermined or ambiguous residues (i.e., “X”, “B” and “Z”), were removed. Finally, all sequences whose 15 C-terminal aa were identical to those of any positive example sequence were excluded. As a result, we obtained a set of 21,028 negative example sequences. Notably, in contrast to previous studies (Boden and Hawkins, 2005; Hawkins et al., 2007), the set of negative examples was not restricted to sequences ending with known PTS1 tripeptides. This way, the training set contains many more examples for learning and on which to model subtle differences in the upstream region.

Since many EST-derived proteins represent close homologs of their query sequence, we also constructed a redundancy-reduced version of the dataset. For this purpose, we clustered the positive and negative example sequences using BLASTclust (http://blast.ncbi.nlm.nih.gov/)

with a similarity threshold of 90% on the 50 C-terminal amino acids of the sequences. Furthermore, sequences with 100% identity in the C-terminal 15 aa were reduced to one representative sequence. The redundancy reduction resulted in 1355 positive and 12512 negative example sequences.

Discriminative machine learning approach. Of the C-terminal 50 aa residues of the example sequences, a maximum of 15 C-terminal residues was considered. Two different models were used to represent sequences and discriminative models: (1) position-specific weight matrices (PWM) and (2) residue interdependence (RI) models. For training of the PWM model, the sequences were represented by an orthonormal encoding method, i.e., each sequence position $i = -15,...,-1$ was represented by a 20-dimensional indicator vector $\mathbf{x}_i \in \{0,1\}^{20}$ with one non-zero entry for the dimension associated with the observed amino acid. Thus, a sequence of length 15 resulted in a 300-dimensional feature vector $\mathbf{x} = [\mathbf{x}_{-15}{}^T,...,\mathbf{x}_{-1}{}^T]^T$ of stacked indicator vectors. In principle, the order in which upstream residues are added to the models associated with different lengths of the C-terminus may influence the model performance. Therefore, different possibilities of extending the tripeptide were evaluated: adding residues in forward/reverse order starting at position -15/-4, respectively, and using 50 random permutations of hexapeptides sampled from the upstream decapeptide.

For the RI model, the sequences were represented as stacked indicator vectors of correlating residues, i.e., each dimension in the feature space is associated with the co-occurrence of two particular residues at two particular positions within the sequence. For example, the second dimension of the RI model feature space is associated with the co-occurrence of Ala at sequence pos. -15 and Ala at sequence pos. -14. As a second example, the 281th dimension is associated with the co-occurrence of Ala at sequence pos. -15 and Val at sequence position -1. As a result, the interdependencies of residues within the C-terminal region are explicitly modeled. This allows, for instance, an analysis of the mutual influence of the PTS1 tripeptide and its upstream region. Note that, in contrast to earlier machine learning approaches (Boden and Hawkins, 2005; Hawkins et al., 2007; Emanuelsson et al., 2003), we do not consider the aa composition of the sequence. First, it has been shown that the composition has no significant influence on prediction performance (Boden and Hawkins, 2005; Hawkins et al., 2007). Second, by abandoning the composition of the complete sequence, our approach may be used for prediction on very short or incomplete 3'-sequences, e.g., ESTs.

For learning of discriminative models we used so-called regularized least squares classifiers (RLSC, Rifkin, 2003). To take into account the significantly imbalanced number of positive and negative examples, we used a modified “balancing” implementation of RLSC

(Lingner and Meinicke, 2008). Using this implementation, the discriminative model can be efficiently calculated for up to ≈ 20,000 dimensions, which is sufficient for the PWM. In contrast, the RI model yields a high-dimensional feature space, e.g., for a sequence of length 15 the RI model feature space comprises 15*(15-1)/2*400+15*20 = 42,300 dimensions. However, to compute a discriminative RIM, we use a kernel-based version of RLSC (Rifkin, 2003). The feature space discriminant for such a model can be calculated using the linear combination of the learned sequence-specific weights and the corresponding sequence feature vectors. The discriminant in feature space allows us to easily interpret discriminative features in terms of relevant sequence properties, e.g., the occurrence of a particular residue at a particular sequence position. The feature space discriminant can also be used to efficiently predict potential peroxisomal proteins.

To evaluate the influence of the RLSC regularization parameter, we partitioned the datasets into three approximately equally sized folds and performed a 3-fold cross-validation using these sets. Thereby, the data set of positive example sequences was randomly split in such a way that each fold contains at least one example of any of the 42 categories of PTS1 tripeptides. The data set of negative example sequences was split randomly. We used $\lambda = \{10^m | m = -4,...,4\}$ as values for the regularization parameter. After training we calculated a model-specific classification threshold by fitting normal distributions to the score distributions of positive and negative test examples, respectively. The threshold can be inferred from the point of intersection of the estimated densities.

To measure the prediction accuracy of our approach, we calculated the average sensitivity, specificity and harmonic mean over the three folds. The sensitivity and specificity are computed as $sens = TP/(TP+FN)$ and $spec = TP/(TP+FP)$, whereby TP is the number of true positives, i.e., the number of positive test examples that score at least as high as the classification threshold. Likewise, FN and FP denote false negatives and false positives, respectively. The harmonic mean is computed as $harm = (2*sens*spec)/(sens+spec)$. The final models were trained with all example sequences using the best regularization parameter found by cross-validation (λ=0.0001 and λ=1000 for the PWM and RI model, respectively). The model-specific classification threshold was calculated using the prediction scores of all example sequences.

Paper II

ARTICLE ADDENDUM

Plant Signaling & Behavior 5:2, 171-175; February 2010; © 2010 Landes Bioscience

Arabidopsis glutathione reductase 1 is dually targeted to peroxisomes and the cytosol

Amr R.A. Kataya and Sigrun Reumann*

Centre for Organelle Research; University of Stavanger; Stavanger, Norway

We recently established a proteome methodology for Arabidopsis leaf peroxisomes and identified more than 90 putative novel proteins of the organelle. These proteins included glutathione reductase isoform 1 (GR1), a major enzyme of the antioxidative defense system that was previously reported to be cytosolic. In this follow-up study, we validated the proteome data by analyzing the in vivo subcellular targeting of GR1 and the function of its C-terminal tripeptide, TNL>, as a putative novel peroxisome targeting signal type 1 (PTS1). The full-length protein was targeted to peroxisomes in onion epidermal cells when fused N-terminally with the reporter protein. The efficiency of peroxisome targeting, however, was weak upon expression from a strong promoter, consistent with the idea that the enzyme is dually targeted to peroxisomes and the cytosol in vivo. The reporter protein that was extended C-terminally by 10 amino acid residues of GR1 was directed to peroxisomes, characterizing TNL> as a novel PTS1. The data thus identify plant peroxisomal GR at the molecular level in the first plant species and complete the plant peroxisomal ascorbate-glutathione cycle. Moreover, GR1 is the first plant protein that is dually targeted to peroxisomes and the cytosol. The evolutionary origin and regulatory mechanisms of dual targeting are discussed.

Key words: ascorbate-glutathione cycle, dual targeting, proteome analyses, reactive oxygen species, targeting signals

Abbreviations: aa, amino acid; APX, ascorbate peroxidase; CaMV, cauliflower mosaic virus; CFP, cyan fluorescent protein; DHAR, dehydroascorbate reductase; EST, expressed sequence tag; GR, glutathione reductase; MDAR, monodehydroascorbate reductase; PTD, peroxisome targeting domain; PTS1/2, peroxisome targeting signal type 1/2; YFP, yellow fluorescent protein

Submitted: 11/04/09

Accepted: 11/05/09

Previously published online: www.landesbioscience.com/journals/psb/article/10527

*Correspondence to: Sigrun Reumann; Email: sigrun.reumann@uis.no

Addendum to: Reumann S, Babujee L, Ma C, Wienkoop S, Siemsen T, Antonicelli GE, et al. Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell 2007; 19:170–93; PMID: 17951448; DOI: 10.1105/tpc.107.050989.

Massive amounts of hydrogen peroxide (H_2O_2) are produced during photosynthesis in peroxisomes by glycolate oxidase activity as part of the photorespiratory cycle.[1] Next to catalase, the ascorbate-glutathione cycle is the secondary scavenging system for H_2O_2 detoxification.[2-4] The cycle comprises four enzymes, ascorbate peroxidase (APX), monodehydroascorbate reductase (MDAR), dehydroascorbate reductase (DHAR) and NADPH-dependent glutathione reductase (GR). GR plays a major physiological role in maintaining and regenerating reduced glutathione in response to biotic and abiotic stresses in plants.[5] Jiminez et al. (1997) provided biochemical evidence for the presence of the antioxidants ascorbate and glutathione and the enzymes of the ascorbate-glutathione cycle in pea peroxisomes.[6-8] While Arabidopsis APX3, MDAR1 and MDAR4 have been characterized as peroxisomal isoforms,[9-11] the molecular identity of plant peroxisomal GR and DHAR have not been determined in any plant species to date.[5] Arabidopsis encodes two GR and five DHAR isoforms that are either shown to be or predicted to be cytosolic, mitochondrial or plastidic.[12] We recently identified specific isoforms of GR (GR1, At3g24170) and DHAR (DHAR1, At1g19570) as being peroxisome-associated by proteome analysis of Arabidopsis leaf peroxisomes.[13,14] Both isoforms were previously reported to be or predicted to be cytosolic.[15]

Arabidopsis GR1 terminates with TNL>, which is related to functional plant PTS1 tripeptides such as SNL> and ANL>.[16,17] Threonine (T), however, has not yet been described as an allowed residue at position -3 of PTS1s in any plant peroxisomal protein.[16] Analysis of homologous plant proteins and expressed sequence tags (ESTs) shows that TNL> is generally highly conserved in putative plant GR1

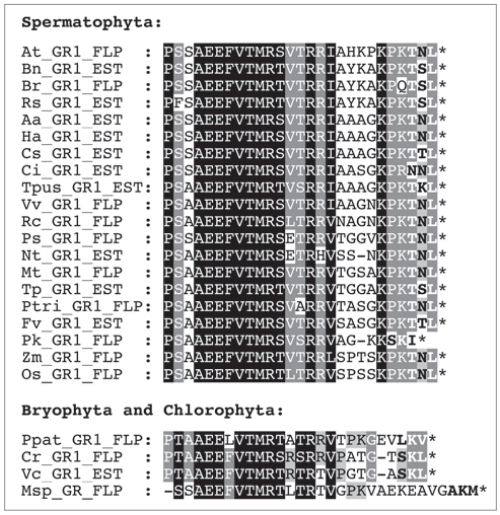


Figure 1. Analysis of PTS1 conservation in plant GR1 homologs. Sequences of full-length protein (FLP) plant GR1 homologs or ESTs ("EST") were identified by BLAST and phylogenetic analysis, aligned by ClustalX, and conserved residues were shaded by Genedoc. In addition to spermatophyta, homologs from bryophyta and chlorophyta were analyzed for PTS1 conservation. For a phylogenetic analysis of the full-length proteins, see also Supplementary Figure 1. The species abbreviations are as follows: Aa, *Artemisia annua*; At, *Arabidopsis thaliana*; Bn, *Brassica napus*; Br, *Brassica rapa*; Ci, *Cichorium intybus*; Cr, *Chlamydomonas reinhardtii*; Cs, *Cynara scolymus*; Fv, *Fragaria vesca*; Ha, *Helianthus annuus*; Msp, Micromonas sp. RCC299; Mt, *Medicago truncatula*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Pk, *Picrorhiza kurrooa*; Ppat, *Physcomitrella patens* subsp. patens; Ps, *Pisum sativum*; Ptri, *Populus trichocarpa*; Rc, *Ricinus communis*; Rs, *Raphanus sativus*; Tp, *Trifolium pratense*; Tpus, *Triphysaria pusilla*; Vc, *Volvox carteri* f. nagariensis; Vv, *Vitis vinifera*; Zm, *Zea mays*.

orthologs (**Fig. 1**). A few other sequences terminate with related tripeptides, such TSL>, TTL>, NNL> and TKL>. Only a single EST (*Picrorhiza kurrooa*) carries the canonical PTS1, SKI> (**Fig. 1**). The data provide only weak additional support for peroxisome targeting of plant GR1 orthologs. However, GR homologs from green algae (chlorophyta) carry canonical PTS1 tripeptides, such as SKL> (Chlamydomonas, Volvox) and AKM> (Micromonas, **Fig. 1**, **Suppl. Fig. 1**).

Arabidopsis GR1 is Targeted to Peroxisomes

To investigate peroxisomal targeting of GR1 by TNL> in vivo, we fused GR1 N-terminally with the reporter protein enhanced yellow fluorescent protein (EYFP) and expressed the construct transiently from the 35S cauliflower mosaic virus (CaMV) promoter in onion epidermal cells and tobacco protoplasts. In parallel, onions were routinely transformed with appropriate negative control plasmids, such as EYFP alone (**Fig. 2A and B**), to verify the absence of endogenous bacteria that may phagocytose cytosolic GFP variants. Although they differed slightly in their active modes of movement, these GFP-labeled subcellular bacteria could, particularly in cases of weak targeting signals and efficiency, be misinterpreted as peroxisomes and cytosolic reporter protein fusions could be falsely identified as being peroxisomally targeted. The N-terminal full-length reporter protein fusion of GR1 (EYFP-GR1) was indeed targeted to small subcellular organelles (**Fig. 2C**), but the efficiency of organelle targeting was weak, making characterization of the structures as peroxisomes difficult in double transformants. In a few cells that co-expressed a peroxisome marker (gMDH-CFP), EYFP-GR1 could conclusively shown to be peroxisomal. The data establish GR1 as a novel peroxisomal enzyme (**Fig. 2C**). In tobacco protoplasts of leaf mesophyll cells, however, EYFP-GR1 remained exclusively cytosolic (**Fig. 2F**), strengthening the idea that negative organelle targeting data may be related to the nature of the analytical in vivo system and require validation in an alternative plant expression system. Notably, the tobacco GR1 ortholog that has been assembled from overlapping ESTs also terminates with TNL> (**Fig. 1**, **Suppl. Fig. 1**).

TNL> is a Novel Plant PTS1

To investigate whether AtGR1 is targeted to peroxisomes by TNL>, we fused the putative peroxisome targeting domain (PTD) comprising the C-terminal ten amino acid (aa) residues of GR1 to EYFP. The extended reporter protein was directed to small subcellular organelles that coincided with peroxisomes (**Fig. 2D**). The peroxisome targeting efficiency ranged from weak to moderate. The same reporter construct also localized to peroxisomes in tobacco mesophyll protoplasts (**Fig. 2E1-5**). In both plant expression systems, the efficiency at which the C-terminal 10 aa residues of GR1 targeted EYFP to peroxisomes was considerably higher compared to full-length GR1. A possible explanation is that the full-length GR1 fusion protein is much larger than the C-terminal 10 aa residues. Due to the C-terminal location of the PTS1 sequence, incomplete transcription and translation products of the gene fusion could remain cytosolic. Alternatively, both EYFP constructs may differ in PTS1 accessibility.

The identification of TNL> as a functional PTS1 tripeptide is important for plant PTS1 protein prediction. First, seven additional Arabidopsis proteins terminating with TNL> emerge as candidate peroxisomal proteins (**Suppl. Table 1**). Second, the data establish T as a new allowed residue in plant PTS1 tripeptides. Six further tripeptides carrying T at pos. -3 (e.g., T[RK][LMI]>) are likely to represent additional functional PTS1 tripeptides. Third, C-terminal tripeptides that are found in a significant number of homologs of established plant PTS1 proteins are considered to be functional plant PTS1 tripeptides. Plant orthologs of GR1 carry a number of non-canonical

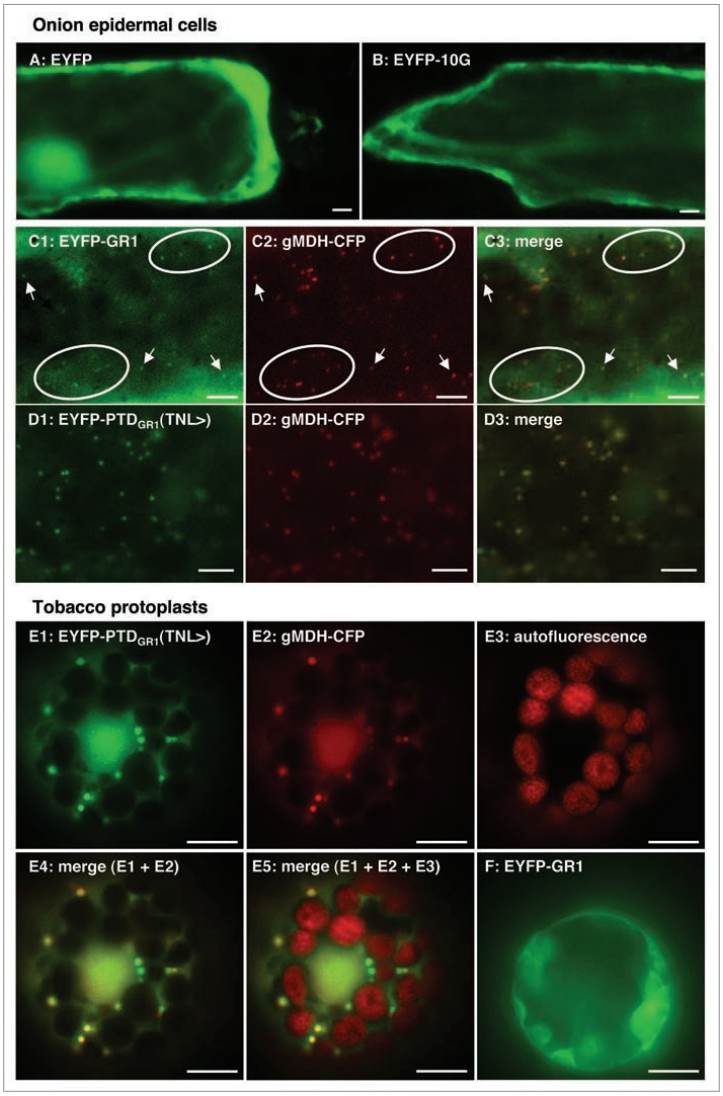


Figure 2. In vivo subcellular targeting analysis of Arabidopsis GR1. The full-length cDNA of Arabidopsis GR1 was fused N-terminally with the reporter protein EYFP and expressed transiently in onion epidermal cells upon biolistic bombardment or in tobacco protoplasts upon polyethylene glycol-mediated transformation. To characterize TNL> as the PTS1 of GR1, EYFP was extended C-terminally by the predicted peroxisome targeting domain of GR1, comprising the C-terminal 10 aa residues (AHKPKPKTNL>). In double transformants, peroxisomes were labeled with gMDH-CFP,[28] and the cyan fluorescence was converted to red to facilitate the detection of green and cyan fluorescent peroxisomes as yellow organelles in image overlays (merge). EYFP alone and EYFP extended C-terminally by 10 glycine residues (EYFP-10G) served as negative controls to verify the absence of endogenous bacteria. Scale bar: 10 μm.

C-terminal tripeptides, such as TSL> (**Fig. 1**), which are likely to emerge as additional plant PTS1 tripeptides in the near future.

Questions arise as to why the PTS1 TNL> has not been detected previously in plant homologs of PTS1 proteins,[16] and why it appears to be restricted to GR1 and possibly a few other proteins. First, the relatively weak targeting efficiency of EYFP-GR1 indicates that the PTS1 domain of GR1 is likely to be insufficient for quantitative peroxisomal targeting of high-abundance peroxisomal matrix proteins. Second, TNL> alone is predicted to be an extremely weak PTS1 tripeptide, requiring auxiliary targeting enhancing elements (e.g., K, P) upstream of the tripeptide (AHKPKPKTNL>) for moderate peroxisome targeting efficiency.[16,18] Most single aa residue mutations, for instance, SNL> to TNL>, are thus likely to be insufficient to maintain peroxisome targeting if they are not paralleled by the conversion of upstream residues to those that enhance targeting. Such multiple mutations, however, are unlikely to occur simultaneously during protein evolution.

Dual Targeting of GR1 to Peroxisomes and the Cytosol

Together with identification of the peroxisomal isoform of DHAR (DHAR1, At1g19570,[14]), all four members of the peroxisomal ascorbate-glutathione cycle have been identified in Arabidopsis at the molecular level. The pronounced conservation of TNL> in higher plant orthologs (**Fig. 1, Suppl. Fig. 1**) strongly suggests that most plant GR1 orthologs, even though they are frequently annotated as cytosolic, are generally dually targeted to peroxisomes and the cytosol. Plant GR1 can thus regenerate reduced glutathione in the peroxisome matrix and substantially contribute to H_2O_2 detoxification. Moreover, peroxisomal GR1 provides the substrate for several other glutathione-dependent enzymes that have been recently discovered in plant peroxisomes.[13,14,19] Glutathione is thereby emerging as a major antioxidant in plant peroxisomes, protecting peroxisomal and cellular enzymes against a range of peroxides, xenobiotics and possibly heavy metals.[20]

Dual or even multiple subcellular targeting of a single protein to several compartments emerges as a more prevalent phenomenon than previously assumed.

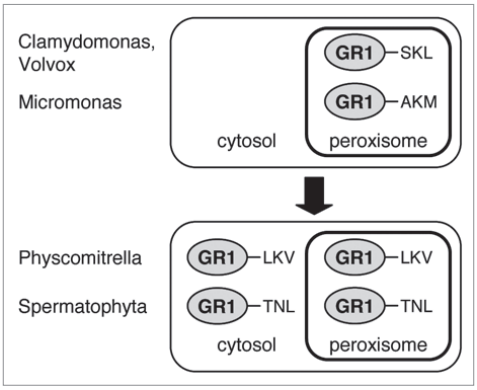


Figure 3. Model of the evolutionary development of the peroxisomal and cytosolic functions of GR1 in spermatophyta. The canonical nature and high peroxisome targeting strength of PTS1s in GR homologs of chlorophyta (Chlamydomonas, Volvox and Micromonas) indicate an exclusive compartmentalization of GR1 in peroxisomes in green algae. By contrast, the non-canonical nature and predicted (LKV>) and proven (TNL>) low peroxisome targeting strength in Physcomitrella and higher plants, respectively, indicate dual enzyme targeting to peroxisomes and the cytosol in bryophyta and spermatophyta.

Six Arabidopsis proteins are dually targeted to mitochondria and peroxisomes,[21,22] and three proteins are dually targeted to plastids and peroxisomes (watermelon Hsp70;[23] soybean aspartate aminotransferase;[24] Arabidopsis 6-phosphogluconolactonase[13]). GR1 is thus the first plant protein that is dually targeted to peroxisomes and the cytosol. In the green lineage, most species bear two GR genes as a result of an early gene duplication event. The two genes cluster into a clade of chloroplastic/mitochondrial isoforms with N-terminal extensions and a clade of cytosolic/peroxisomal isoforms terminating with known or PTS1-related tripeptides (**Suppl. Fig. 1**). PTS1 conservation analysis in cytosolic/peroxisomal isoforms allows the prediction as to whether the cytosolic or peroxisomal function of GR1 is the most fundamental and most evolutionarily ancient task. As revealed by phylogenetic analysis of full-length protein sequences and cDNAs assembled from ESTs, one of two GR homologs from green algae carries a strong canonical PTS1 tripeptide, such as SKL> (Chlamydomonas, Volvox) and AKM> (Micromonas). The data indicate that the peroxisome-targeting function of GR evolved soon after gene duplication in Viridiplantae and that these GR1 homologs are most likely exclusively peroxisomally targeted in chlorophyta (**Suppl. Fig. 1**, see also **Fig. 3**). Dual targeting of GR1 to peroxisomes and the cytosol appears to have evolved at a later stage, prior to the divergence of mosses. The GR1 homolog from Physcomitrella carries a PTS1-related tripeptide, LKV>, of weak predicted targeting strength, similar to TNL> in higher plants. Notably, leucine (L) has recently been experimentally verified as another allowed residue at pos. -3 in plant PTS1 tripeptides (Lingner T, Antonicelli GE, Kataya A, Reumann S, unpubl. data).

The most intriguing question of multiple protein targeting concerns the possible existence and identity of the underlying regulatory mechanism. Two mechanisms that regulate protein distribution between the cytosol and peroxisomes can be envisioned: (1) a rather static mechanism of protein distribution that is largely determined by the PTS strength, i.e., the affinity of the PTS1 domain to its receptor, PEX5p, and (2) a dynamic mechanism of post-transcriptional or post-translational nature that allows adequate responses to changing demands in GR1 compartmentalization. Notably, even in the case of a rather simple regulation of protein distribution by PTS1 strength, the peroxisome import rate by weak PTS1s in vivo depends on the rate of gene expression; low-abundance proteins with weak PTS1s may still be imported quantitatively into peroxisomes, but not high-abundance proteins. Therefore, the low peroxisome import rate of GR1 upon expression from the strong 35S promoter does not allow the conclusion to be made that GR1 is largely cytosolic in vivo. This question needs to be addressed by expressing N-terminally or internally tagged GR1 versions, preferentially in a *gr1* knockout background, from the native GR1 promoter. The fact that GR1 was identified in leaf peroxisomes that were isolated from plants grown under standard conditions[13,14] supports the idea of a constitutive role of GR1 in peroxisomes.

In light of the strong need and major function of the peroxisomal ascorbate-glutathione cycle under sudden conditions of catalase inactivation,[25] it is reasonable to postulate that the ratio of GR1 distribution between peroxisomes and the cytosol is dynamic and adjustable depending on H_2O_2 overproduction in the matrix by post-transcriptional and/or post-translational mechanisms. Interestingly, pea (and later, Arabidopsis) GR2 was one of the first proteins shown to be dually targeted to two plant cell compartments, mitochondria and plastids.[12,26] Dual targeting of GR1/2 thus emerges as a specific property of the enzyme. The regulatory mechanism(s) underlying dual targeting of both GR1 and GR2 are still unknown and remain to be studied. The exclusively cytosolic localization of EYFP-GR1 in tobacco protoplasts (**Fig. 2**) may indicate that peroxisome targeting is prevented in this expression system by PTS1 inaccessibility caused by conformational changes or binding of an accessory protein that regulates enzyme distribution between the cytosol and peroxisomes.

Conclusions

The data presented here conclusively identify the peroxisomal GR isoform in the first plant species at the molecular level, thereby completing the ascorbate-glutathione cycle in plant peroxisomes. In vivo studies, for instance, reverse genetics, as have been recently initiated for cytosolic *gr1*,[27] need to follow in order to investigate the functions of cytosolic and peroxisomal GR1 in planta and "*in organello*." Future studies will also need to focus on the regulatory mechanism of dual targeting.

Acknowledgements

We would like to thank the Arabidopsis Biological Resource Center (ABRC) for providing the GR1 full-length cDNA and Aline Benichou and Dr. Xiong-Yan Chen for experimental support.

Note

Supplementary materials can be found at: www.landesbioscience.com/supplement/KatayaPSB5-2-Sup.pdf

References

1. Huang AHC, Moore TS, Trelease RN. Plant peroxisomes. New York: Academic Press 1983.
2. Halliwell B, Foyer CH. Ascorbic acid, metal ions and the superoxide radical. Biochem J 1976; 155:697-700.
3. del Rio LA, Sandalio LM, Corpas FJ, Palma JM, Barroso JB. Reactive oxygen species and reactive nitrogen species in peroxisomes. Production, scavenging and role in cell signaling. Plant Physiol 2006; 141:330-5.
4. Noctor G, Foyer CH. ASCORBATE AND GLUTATHIONE: Keeping Active Oxygen Under Control. Annu Rev Plant Physiol Plant Mol Biol 1998; 49:249-79.
5. Romero-Puertas MC, Corpas FJ, Sandalio LM, Leterrier M, Rodriguez-Serrano M, Del Rio LA, et al. Glutathione reductase from pea leaves: response to abiotic stress and characterization of the peroxisomal isozyme. New Phytol 2006; 170:43-52.
6. Jimenez A, Hernandez JA, Del Rio LA, Sevilla F. Evidence for the presence of the ascorbate-glutathione cycle in mitochondria and peroxisomes of pea leaves. Plant Physiol 1997; 114:275-84.
7. Jimenez A, Hernandez JA, Pastori G, del Rio LA, Sevilla F. Role of the ascorbate-glutathione cycle of mitochondria and peroxisomes in the senescence of pea leaves. Plant Physiol 1998; 118:1327-35.
8. Palma JM, Jimenez A, Sandalio LM, Corpas FJ, Lundqvist M, Gomez M, et al. Antioxidative enzymes from chloroplasts, mitochondria and peroxisomes during leaf senescence of nodulated pea plants. J Exp Bot 2006; 57:1747-58.
9. Zhang H, Wang J, Nickel U, Allen RD, Goodman HM. Cloning and expression of an Arabidopsis gene encoding a putative peroxisomal ascorbate peroxidase. Plant Mol Biol 1997; 34:967-71.
10. Leterrier M, Corpas FJ, Barroso JB, Sandalio LM, del Rio LA. Peroxisomal monodehydroascorbate reductase. Genomic clone characterization and functional analysis under environmental stress conditions. Plant Physiol 2005; 138:2111-23.
11. Lisenbee CS, Lingard MJ, Trelease RN. Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. Plant J 2005; 43:900-14.
12. Chew O, Whelan J, Millar AH. Molecular definition of the ascorbate-glutathione cycle in Arabidopsis mitochondria reveals dual targeting of antioxidant defenses in plants. J Biol Chem 2003; 278:46869-77.
13. Reumann S, Babujee L, Ma C, Wienkoop S, Siemsen T, Antonicelli GE, et al. Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways and defense mechanisms. Plant Cell 2007; 19:3170-93.
14. Reumann S, Quan S, Aung K, Yang P, Manandhar-Shrestha K, Holbrook D, et al. In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol 2009; 150:125-43.
15. Stevens RG, Creissen GP, Mullineaux PM. Cloning and characterisation of a cytosolic glutathione reductase cDNA from pea (*Pisum sativum* L.) and its expression in response to stress. Plant Mol Biol 1997; 35:641-54.
16. Reumann S. Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol 2004; 135:783-800.
17. Nowak K, Luniak N, Witt C, Wustefeld Y, Wachter A, Mendel RR, et al. Peroxisomal localization of sulfite oxidase separates it from chloroplast-based sulfur assimilation. Plant Cell Physiol 2004; 45:1889-94.
18. Mullen RT, Lee MS, Flynn CR, Trelease RN. Diverse amino acid residues function within the type 1 peroxisomal targeting signal. Implications for the role of accessory residues upstream of the type 1 peroxisomal targeting signal. Plant Physiol 1997; 115:881-9.
19. Dixon DP, Hawkins T, Hussey PJ, Edwards R. Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. J Exp Bot 2009; 60:1207-18.
20. Noctor G, Gomez L, Vanacker H, Foyer CH. Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. J Exp Bot 2002; 53:1283-304.
21. Carrie C, Murcha MW, Kuehn K, Duncan O, Barthet M, Smith PM, et al. Type II NAD(P)H dehydrogenases are targeted to mitochondria and chloroplasts or peroxisomes in *Arabidopsis thaliana*. FEBS Lett 2008; 582:3073-9.
22. Carrie C, Kuhn K, Murcha MW, Duncan O, Small ID, O'Toole N, et al. Approaches to defining dual-targeted proteins in Arabidopsis. Plant J 2009; 57:1128-39.
23. Wimmer B, Lottspeich F, van der Klei I, Veenhuis M, Gietl C. The glyoxysomal and plastid molecular chaperones (70-kDa heat shock protein) of watermelon cotyledons are encoded by a single gene. Proc Natl Acad Sci USA 1997; 94:13624-9.
24. Gebhardt JS, Wadsworth GJ, Matthews BF. Characterization of a single soybean cDNA encoding cytosolic and glyoxysomal isozymes of aspartate aminotransferase. Plant Mol Biol 1998; 37:99-108.
25. Feierabend J, Engel S. Photoinactivation of catalase in vitro and in leaves. Arch Biochem Biophys 1986; 251:567-76.
26. Creissen G, Reynolds H, Xue Y, Mullineaux P. Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. Plant J 1995; 8:167-75.
27. Marty L, Siala W, Schwarzlander M, Fricker MD, Wirtz M, Sweetlove LJ, et al. The NADPH-dependent thioredoxin system constitutes a functional backup for cytosolic glutathione reductase in Arabidopsis. Proc Natl Acad Sci USA 2009; 106:9109-14.
28. Ma C, Haslbeck M, Babujee L, Jahn O, Reumann S. Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. Plant Physiol 2006; 141:47-60.