Identification and characterization of putative Toc159 interacting partners

A dissertation submitted to the Faculté des Sciences
Institut de Biologie
Université de Neuchâtel

For the degree of
Docteur ès Sciences+

by

Cyrille Montandon

Accepted after recommendation by the thesis committee:

Prof. Felix Kessler, Thesis Director
Prof. Michel Goldschmidt-Clermont
Prof. Jean-Marc Neuhaus
Dr Birgit Agne

Defended on May 28th, 2015

Université de Neuchâtel
2015

UNIVERSITÉ DE NEUCHÂTEL

Faculté des sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
E-mail: secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Cyrille MONTANDON

Titre:

"Studies on the role of an LRR (Leucine Rich Repeat) protein in chloroplast protein import, in A. thaliana"

sur le rapport des membres du jury composé comme suit:

- Prof. Felix Kessler, directeur de thèse, Université de Neuchâtel
- Prof. Jean-Marc Neuhaus, Université de Neuchâtel
- Prof. Michel Goldschmidt-Clermont, Université de Genève
- PD Dr Christian Parisod, Université de Neuchâtel
- Dr Birgit Agne, Martin-Luther Universität, Halle, Allemagne

Neuchâtel, le 26 juin 2015

Le Doyen, Prof. B. Colbois

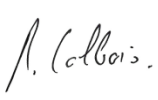


Table of content:

Abstract : 9
Résumé: 10
Introduction: 11
The Endosymbiotic theory of organelles origins (fig. 1): 11
Chloroplast structures and functions (fig. 2, a-b): 12
Import pathways (fig. 2, c): 13
Specific Protein sorting to the organelle: 14
Chloroplast protein import: 15
In vitro import of a pre-protein in the chloroplast: 15
Role of ATP in the different stage of the import: 16
A general import pathway? 16
Identification of the putative receptors: 17
Sequence of the pre-protein interactions with the Toc/Tic proteins: 18
Biochemical vs genetic studies: 19
Toc/Tic components: 19
Toc86-Toc159: 19
Toc34/Toc33: 21
Toc75: 21
Tic110: 22
Tic40: 22
Tic20 and Tic22: 23
Tic21: 23
Chaperones and molecular motor: 23
Stromal processing peptidase: 24
Toc/Tic complexes: 24
The Tic 1 MD complex : 25
Role of the Toc159 and Toc33 GTPase in protein import: 25
Small GTPase: 25
Toc33 homodimer and Toc33/Toc159 hetrodimer: 26
Effect of Toc159 GTPase mutation in vivo: 26
Model of the Toc/Tic protein import process (fig. 4): 27
Evolution of the import apparatus: 27
Transit peptide, a hidden specificity: 27

Selectivity and regulation of the import activity: 28
Developmental and environmental variations of the import efficiency: 28
Different plastid types have overlapping but different import specificity: 29
Specialized import unit in *A. thaliana*: 29
Post-translational regulation of the import: 30
A role for A-domain post-translational modification? 31
Aim of the thesis: 32
Material and methods : 35
Material: 35
Plants lines: 35
Bacteria strains: 35
Yeast strains: 35
Chemicals: 35
Plasmids: 35
Antibodies: 36
Methods: 36
Plant methods: 36
Chloroplasts isolation: 37
In vitro import: 38
TAP-Toc159 purifications : 39
Stable and transient transformation: 40
Proteins methods: 41
Yeast: 41
Yeast media 41
DNA cloning, genotyping TAIL PCR and qPCR: 43
Microscopy: 47
Mass-spectrometry: 47
Bioinformatic: 48
Results: 49
TAP-Toc159 purification for MS analysis: 49
Purification and western blotting: 49
Mass-spectrometry (MS) analysis and filtering of the proteins identified: 51
AT4G322250, a predicted protein kinase at the chloroplast envelope: 54

Bioinformatic data (fig. 10): 54
In vitro import of AT4G322250 (Fig. 11): 55
Transient expression AT4G322250-YFP in tobacco leaves: 56
Co-purification of AT4G322250 with TAP-Toc159 (Fig. 13): 56
AT5G01590-Tic56: 59
Emb2004-AT1G10510: 63
Emb2004, a LRR protein (fig. 19 and fig. 20): 63
In vitro import of Emb2004 and proteases protection assay (fig. 21): 65
Emb2004-YFP localization (fig: 22): 67
Sub-chloroplastic localization of Emb2004 (Fig. 23): 67
Antibody production (Fig. 25): 69
Emb2004 was not detected in a TAP-Toc159 using the anti-Emb2004 antibody (fig. 25): 70
Yeast two hybrid assay (Fig. 26): 72
Emb2004-TAP line (Fig. 28): 73
TAIL analysis of the T-DNA insertion in the Emb2004-Tap plant line (Fig. 29): 76
Analysis of the silenced line (Fig. 30-31): 77
Complementation of the *emb2004* mutant (Fig. 32-33): 81
Purification of the Emb2004-TAP protein (Fig. 39): 82
Identification of putative interacting partners of Emb2004: 83
Discussion: 93
AT4G32250 kinase: 93
AT5G01590/Tic56: 94
Emb2004: 95
Localisation: 95
Failure of Emb2004 to interact with Toc159: 96
Complementation of the emb2004 mutant by Emb2004-TAP 97
Pale phenotypes and undeveloped chloroplast observed in Emb2004-TAP lines: 97
Purification of Emb2004-TAP and MS analysis: 98
Yeast two hybrids screens: 98
Conclusion: 99
References: 101
Collaborations: 115
Mass-spectrometry: 115

AT4G32250 project: 115
AT5G01590/Tic56 project: 115
Emb2004/At1G10510 project: 115
Papers: 115
Acknowledgements: 117
Annex: 118

Abstract :

The chloroplast is the distinctive organelle of land plants and other photosynthetic eukaryotes. It carries out a variety of metabolic process, but photosynthesis is its main task and this is reflected in its structures and protein composition. Chloroplasts evolved from an endosymbiosis between a photosynthetic cyanobacterium and an ancestral eukaryote. Reminiscent of this autonomous origin, the chloroplast genome encodes approximately 100 proteins. The majority of the remaining 2000-3000 proteins identified in the chloroplast are encoded in the nucleus and translated in the cytosol and must be imported into the chloroplast. The transit peptide, a small sequence at the N-terminus of the proteins destined for the chloroplast, is necessary and sufficient for specific import into the chloroplast. The Toc/Tic (Translocon at the Outer/Inner membrane of the chloroplast envelope) pathway mediates the recognition and the translocation of these proteins in an ATP- and GTP-dependent way. The Toc core complex is constituted of two receptors with a GTPase domain, Toc159 and Toc34/33, and a channel, Toc75. Toc159 is formed of three domains, a C-terminal membrane anchoring domain (M domain), a GTPase domain (G domain) and an acidic domain (A-domain). The protein import rate and specificity toward client proteins vary depending on the developmental stage, tissue or plastid type. The different *A. thaliana* homologues of Toc159 and Toc34/33 form distinct complexes. They have different import specificities and their expression level depends on the developmental stage and the anatomical part of the plant. The A-domain of Toc159 and its homologues partially determine the specificity toward the different types of client proteins. The A-domain is hyper-phosphorylated and exists as a soluble form, likely the result of a specific cleavage. These post-translational modifications of the A-domain might play a role in determining the affinity of Toc159 toward client proteins. The control of the GTPase activity of Toc159 or Toc34/33 might also have an influence on the import. The aim of this thesis was the characterization of proteins potentially responsible for these modifications and that were co-purified with a tagged Toc159 and identified by mass-spectrometry. The main effort was made on Emb2004/AT1G10510, a LRR (Leucine Rich Repeat) protein and contributions were made for the characterization of a predicted protein kinase (AT4G32250) and Tic56 (AT5G01590).

Keywords:

chloroplasts; protein import; translocon; Toc159; Emb2004; Leucin Rich Repeat (LRR); Tic56; AT4G322250; kinase; A-domain; transit peptide;

Résumé:

Le chloroplaste est l'organelle qui caractérise les plantes terrestres ainsi que les autres eucaryotes photosynthétiques. Il remplit diverses fonctions métaboliques, mais la photosynthèse est son activité principale, sa structure et sa composition protéique en témoignent. Le chloroplaste est le résultat d'une endosymbiose entre un eucaryote ancestral et une cyanobactérie photosynthétique. Le génome du chloroplaste, vestige de son ancienne autonomie, encode environ une centaine de protéines. Les 2000-3000 autres protéines présentes dans le chloroplaste sont codées dans le noyau et traduites dans le cytosol et doivent donc être importées dans le chloroplaste. Le « transit peptide » (peptide de transit), une courte séquence à la terminaison aminée des protéines destinées au chloroplaste, est suffisant et nécessaire pour l'import spécifique dans le chloroplaste. La voie Toc/Tic (Translocon at the Outer/Inner membrane of the chloroplast envelope: « Translocon à la membrane externe/interne de l'enveloppe du chloroplaste ») reconnait et import ces protéines en présence d'ATP et de GTP. Le noyau du complexe Toc est composé de Toc159 et Toc34/33, 2 récepteurs possédant un domaine GTPase, et de Toc75 qui constitue le pore du complexe. Toc159 est formé de 3 domaines, un domaine ancrant la protéine dans la membrane à l'extrémité carboxyle (domaine M), un domaine GTPase (domaine G) et un domaine acide (domaine A). Le taux d'import des protéines et sa spécificité envers les protéines clientes varient en fonction du développement de la plante, du tissu ou du type de plaste. Les différents homologues de Toc159 et de Toc34/33 chez *A. thaliana* forment des complexes distincts. Ils montrent une spécificité d'import différente et leur expression varie en fonction du stade de développement ou de la partie de la plante. Le domaine A de Toc159 et de ses homologues détermine partiellement la spécificité envers les différents types de protéines clientes. Le domaine A est hyper-phosphorylé et existe sous une forme soluble, certainement le résultat d'un clivage spécifique. Ces modifications post-traductionnelles pourraient influencer la spécificité de Toc159 envers les protéines clientes. Le contrôle de l'activité GTPase de Toc159 ou de Toc34/33 pourrait également influencer l'activité d'import. Le but de ce travail de thèse était la caractérisation de protéines co-précipitées avec une version taguée de Toc159 et identifiées par spectrométrie de masse, qui pourraient être potentiellement responsable des modifications mentionnées ci-dessus. L'effort principal a été fait sur Emb2004/AT1G10510, une protéine LRR (Leucine Rich Repeat) et des contributions ont été faites à la caractérisation d'une protéine kinase potentielle/AT4G32250 et de Tic56/AT5G01590.

Introduction:

The Endosymbiotic theory of organelles origins (fig. 1):

At the end of the 19th century, the German botanist Andreas Schimper noticed that the chloroplast binary division resembled that of cyanobacteria. 20 years later, Mereschowsky, drawing parallels with the symbiotic cooperation of fungi and algae in lichens, suggested first the ideas that would later evolves into endosymbiosis theory, describing chloroplasts "as green little slaves ". He hypothesized that the ancestor of plants might have engulfed the prokaryotic ancestor of the chloroplast, both developing a mutual dependence through the generations (McFadden 1999). In the second half of the 20th century, experimental evidence came in support of this hypothesis. The incorporation of radioactive thymidine in the *Spirogyra* chloroplast revealed the presence of DNA. Later on, different analyses of the chromosome of the chloroplast concluded that it shares common feature with bacterial DNA. This was confirmed by the first sequences of chloroplast genes in 1978 (Schwartz & Dayhoff 1978).

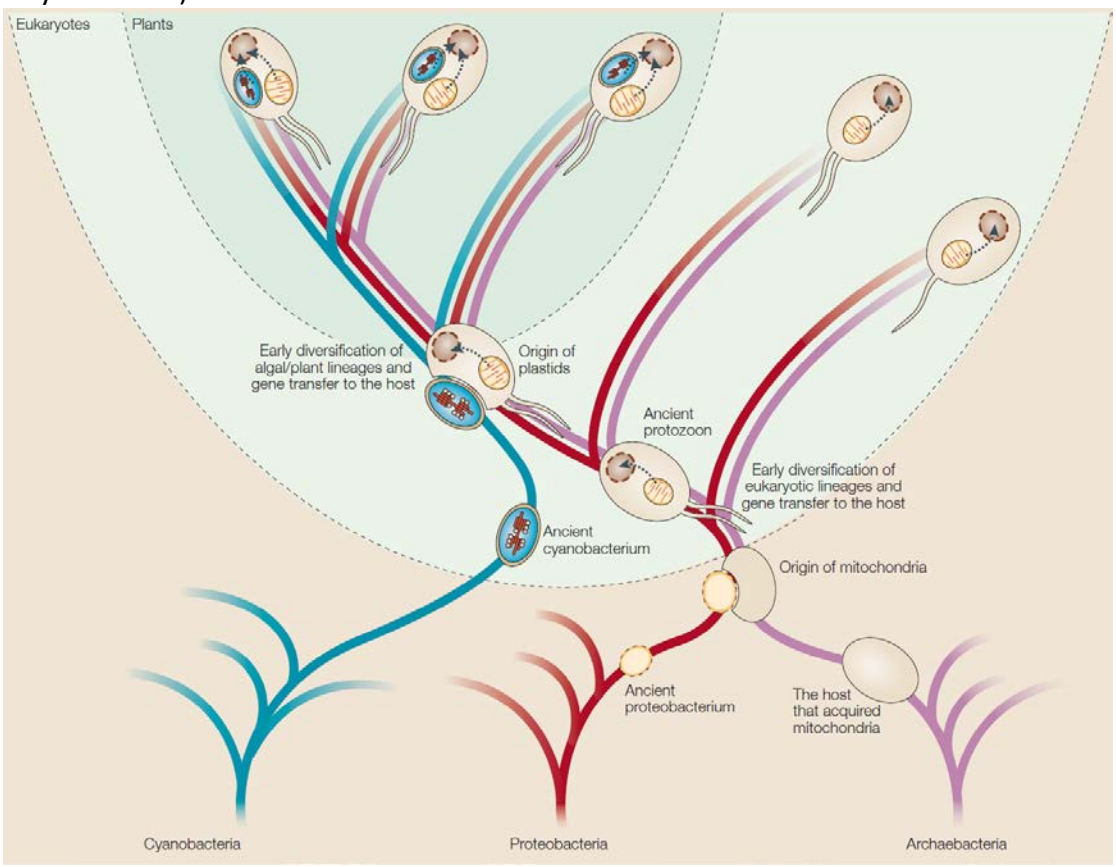


Figure 1: Both mitochondria and chloroplast originate from endosymbiotic events. An ancient proteobacterium is thought to be the ancestor of the mitochondria which was engulfed by an ancient archaebacterium. The chloroplast originates from a later endosymbiotic event between an ancient eukaryote and a cyanobacterium. As for the mitochondria, the endosymbiosis became permanent and most genes of the primitive chloroplast were transferred to the nucleus. The similarities between the division process of the chloroplast and cyanobacteria were the first indication that they might share a common origin. This was later supported by the discovery of the chloroplast genome and its sequencing, demonstrating the prokaryotic origin of chloroplast genes (Timmis et al. 2004).

Lynn Margulis was instrumental in developing and promoting the theory of the endosymbiotic origin of the chloroplast (Sagan 1967). Mitochondria were shown to originate from a similar endosymbiotic process (fig. 1) (Schwartz & Dayhoff 1978).

Chloroplast structures and functions (fig. 2, a-b):

Chloroplasts are the distinctive organelles of *Plantae* and the photosynthetic eukaryotes of other kingdoms. It descended from a cyanobacterium which established an endosymbiosis with an ancient protozoon 1-1.5 billion years ago. The mean chloroplast number in a *A. thaliana* mesophyll cell is 120 (Pyke & Leech 1994) and chloroplasts contain up to 75% of the cellular nitrogen content (Peoples & Dalling 1988). The main task of the chloroplast is photosynthesis, i.e. producing reducing power and chemical energy with light, to convert carbon dioxide and water into carbohydrates and molecular oxygen. The chloroplast is enclosed by a double membrane envelope, a remnant of its endosymbiotic origin, enclosing the soluble intermembrane compartment (IMS). The soluble compartment enclosed by the envelope is called the stroma and contains mainly the enzymes responsible for the light-independent carbon dioxide fixation. The photosynthetic proteins and pigments reside on the thylakoids membrane, an interconnected membranous compartment organized in stacks called grana (fig. 2, a- b). The light-dependent reactions on the thylakoid membrane induce a pH gradient between the stroma and the thylakoid lumen, which is used for the production of ATP. The photosynthetic proteins quantitatively represent the majority of the chloroplast proteins, Rubisco accounting for more than 50% of the soluble protein (Spreitzer & Salvucci 2002).

Beside photosynthesis, chloroplasts are involved in many other important metabolic tasks, like synthesis of amino-acids, fatty acids, pyrimidine bases, terpenoid and nitrogen assimilation. They also take part in the synthesis of secondary metabolites such as pigments and hormones (Jarvis & López-Juez 2013). The originally photosynthetic chloroplast diversified into other types of plastid with different function. The amyloplasts are mainly present in roots and are involved in the storage of large quantities of starch and required for gravitropism. The chromoplasts have a more ecological function, like in fruit for attracting animals for seed dispersal. Etioplasts are present in photosynthetic tissue when the plant is grown in the dark. They develop into chloroplasts once exposed to light. All types of plastids differentiate from the same proplastids that are present in the meristematic tissues. Plastids have no fixed status and can evolve into another type of plastid under particular developmental or environmental condition (Bob Buchanan, Wilhelm Gruissem 2002). For instance mature chloroplast de-differentiate into proplastids in artificial cell culture condition (Harikrishna et al. 1992). Chloroplasts have their own genome, hence their own transcription and translation machinery, but the majority of the chloroplast proteins are encoded in the nucleus and transcribed in the cytosol. Some of the metabolic tasks described above can involve protein components of both nuclear and chloroplast origin, sometimes even in the same protein complex. The chloroplast evolved a complex protein machinery to specifically import its protein to its different compartments (Jarvis & López-Juez 2013).

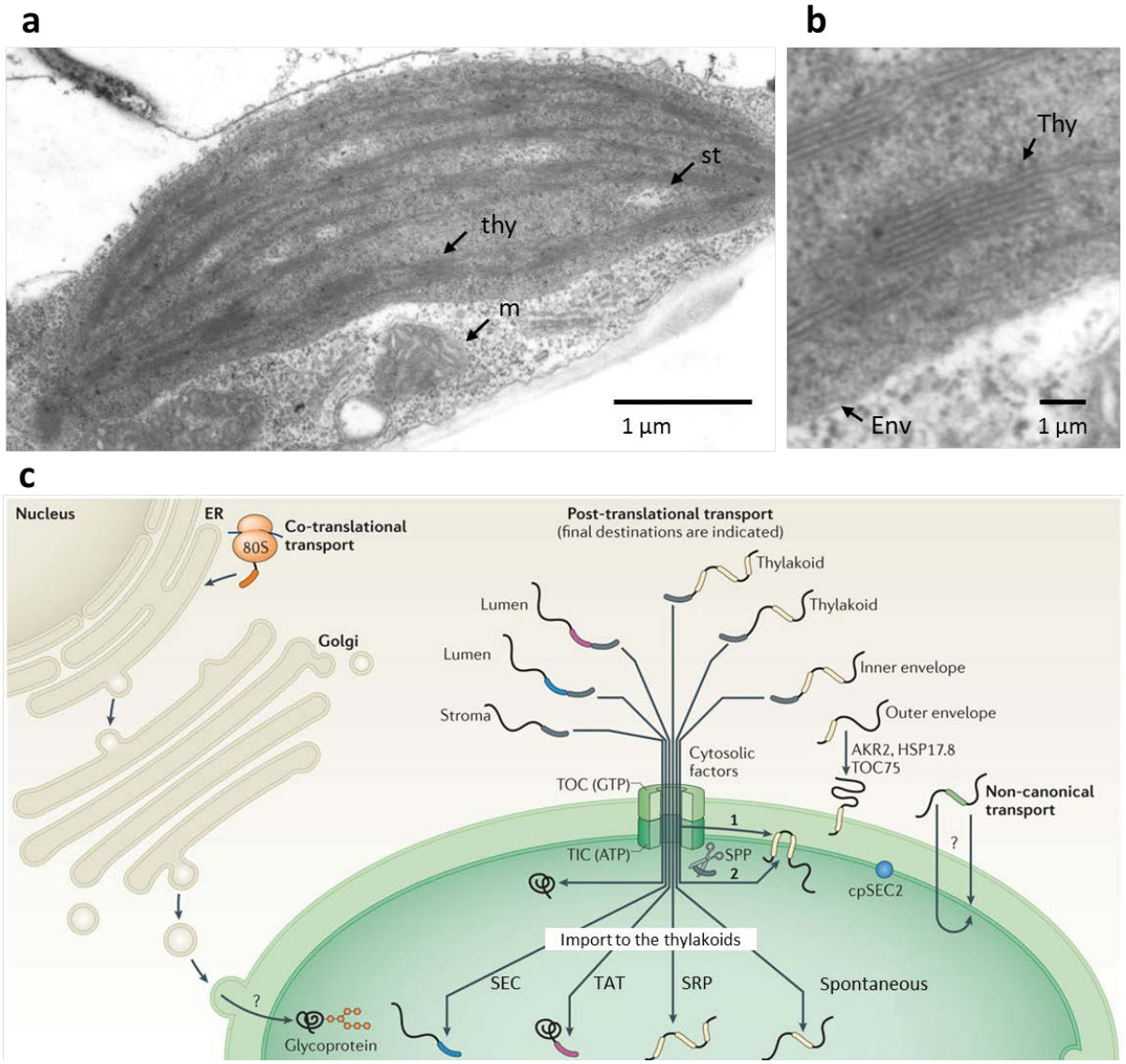


Figure 2: The Toc/Tic pathway is the main translocation pathway at the envelope of the chloroplast: a) transmission electron micrograph of *A. thaliana* chloroplast ultrastructure. thy: thylakoid membranes; st. starch granule. m: mitochondria. **b)** Magnification of micrograph **a.** Thy: thylakoid grana stacks. Env: envelope composed of the outer and the inner membrane. **c)** The Toc/Tic pathway mediates the specific import of the proteins into the chloroplast through the recognition of a transit peptide. Outer membrane protein insertion is mediated by cytosolic AKR2 and HSP17.8 factor and Toc75 outer envelope protein. A non-canonical transport pathway is thought to import chloroplast proteins lacking a transit peptide to the inner membrane of the envelope. Some proteins are co-translationally imported in the ER and targeted to the Golgi apparatus and then the chloroplast envelope, through vesicular transport. Inside the chloroplast, three distinct routes, the SEC, TAT and SRP pathways, mediate the targeting of the proteins to the thylakoid membrane or lumen. Some proteins insert spontaneously in the thylakoid membranes. Adapted from (Jarvis & López-Juez 2013).

Import pathways (fig. 2, c):

The Toc/Tic (Translocon at the Outer/Inner membrane of Chloroplast) pathway is the main route of protein import into the chloroplast. It imports proteins through the specific recognition of a N-terminal transit peptide which is cleaved upon arrival in the stroma. The Toc/Tic machinery mediates the import of protein to the intermembrane space, the inner

membrane or the stroma. Toc75 is so far the only known outer envelope protein imported through the Toc/Tic pathway. The Toc/Tic import pathway shares convergent features with other organellar targeting systems, such as the presence of a cleavable targeting sequence (the transit peptide), but it has a distinct evolutionary origin (Jarvis & Soll 2001). Most of the chloroplast proteins have a predicted transit peptide (Bruce 2000). Three additional transport routes have been reported. The AKR2 cytosolic sorting factor together with HSP17.8 mediates the insertion of outer envelope proteins (Hofmann & Theg 2005). Some proteins without a cleavable transit peptide are imported through an alternative but poorly studied pathway (Miras et al. 2007; Armbruster et al. 2009). Finally, it appears that some proteins are first directed to the ER and reach the chloroplast via vesicular trafficking (Villarejo et al. 2005). In the chloroplast, three distinct pathways namely the TAT (Twin-Arginine Translocation) (Robinson & Bolhuis 2004), the SEC pathway (Secretory pathway) and SRP (Signal Recognition Particle) pathway mediate the targeting of proteins to the thylakoid membrane and lumen (Albiniak et al. 2012) (fig. 2, c). The rest of this introduction gives a brief historical point of view on the characterization of protein targeting in the cell and more particularly in the chloroplast. In the end, the Toc/Tic pathway, its components and mechanism are discussed in more details.

Specific Protein sorting to the organelle:

As the symbiosis evolved, most of the genes of the chloroplast migrated to the nucleus. Now, of approx. 2000-3000 proteins present in the *A. thaliana* chloroplast, only 100 are still encoded in the chloroplastic genome (Sato et al. 1999). Thus, the plant had to develop a mechanism to direct its proteins which are encoded in the nucleus and synthetized in the cytosol back to the chloroplast. More generally, eukaryotic cells are composed of different compartments enclosed by one or several membranes. As all of these compartments have a specialized function and set of proteins, one has to wonder how proteins synthetized in the cytosol are properly targeted to the right compartment. Different mechanisms were hypothesized, some involving the interaction of an untranslated 5' part of the mRNA with the compartment of destination, where it is translated and the resulting peptide imported. Another hypothesis was that there are different types of ribosome, each specifically interacting with one compartment and type of mRNA.

In 1971, Blobel & Sabatini (1971) proposed the hypothesis that the specificity for an organelle is mediated by the nascent protein chain. They developed a model for protein import called the "signal hypothesis" hypothesizing the existence of the following components: 1) a soluble factor mediating the interaction between the N-terminal peptide chain and the organelle. 2) An aqueous channel, through which the synthetized protein migrates. 3) A protease at the trans-side of the organelle membrane which cleaves off the N-terminal part of the protein upon import. In 1972 it was shown that an IgG light chain synthetized in a cell-free system was longer than the *in vivo* secreted form (Milstein C, Brownlee GG, Harrison TM 1972). In 1975 (G Blobel & Dobberstein 1975b; G Blobel & Dobberstein 1975a) demonstrated that an IgG light chain synthesized on polysomes of the rough ER microsomal fraction had the same length as the in vivo secreted proteins (Gonter Blobel & Dobberstein 1975). This translation product was resistant to protease treatment,

demonstrating that it must be inside the microsomal vesicles. When the polysomes were stripped from their microsomal fraction, they began synthetizing a longer IgG light chain, indicating that the ER microsomes contain a factor able to cleave off the N-terminal part of the newly synthetized protein. They managed to reconstitute the translation and processing of the IgG with heterologous ribosomes and microsomes with IgG mRNA but not with hemoglobin mRNA, accounting for the specificity of the process (Blobel & Dobberstein 1975). Later on, protein factors were shown to be implicated in the process. It appears that the IgG light chain could not be imported post-translationally, but a post-translational route of protein import in the ER of yeast does exist (Johnson et al. 2013). The signal hypothesis of specific targeting of proteins to organelles proved to be correct for the different cell compartments, including mitochondria and chloroplasts. However, unlike the ER/secretory pathway these import processes occur post-translationally. It was later shown that the N-terminal signal peptide is necessary and sufficient for proper protein addressing (Reviewed in Blobel's Nobel lecture: Blobel 2000).

Chloroplast protein import:

This part describes chronologically the main discoveries which finally led to the identification of the components of the chloroplast import machinery. Some key biochemical techniques helped unravelling the chloroplast import process and led to the identification of its components. *In vitro* import of a synthetic pre-protein into isolated chloroplasts is the cornerstone for the understanding of protein import into the chloroplast. It was used to study the energy requirements of the import process, and led to the identification of distinct steps, each of it corresponding to different ATP requirements. It also resulted in the identification of the protein factors involved in import, by co-isolation with a ProteinA-tagged pre-protein engaged in the import process at either “low” or “high” concentrations of ATP and analyzing the co-isolated proteins. Label transfer crosslinking was used to determine which proteins directly interact with different part of the pre-protein and what is the chronological sequence of these interactions.

In vitro import of a pre-protein in the chloroplast:

In 1978, Chua et al, demonstrated that an *in vitro* synthetized SSU was longer than the endogenous one (Chua & Schmidt 1978). They managed to import this longer protein into isolated chloroplast in presence of ATP, where it was processed to a smaller “mature” size identical to the endogenous protein and successfully integrated into the Rubisco holoenzyme. This established the posttranslational nature of chloroplast protein import and the presence of a cleavable sequence. Chloroplastic proteins without their transit peptide - i.e. the N-terminal part which is cleaved upon arrival in the stroma- cannot be imported (Mishkind et al. 1985) whereas non-chloroplastic protein fused to a transit peptide are successfully translocated, hence the transit peptide in necessary and sufficient for a protein to be imported in the chloroplast (Lubben et al. 1988; Friedman & Keegstra 1989).

Role of ATP in the different stages of the import:

Cline et al. established the ATP-independent binding of the pre-protein to the surface of the chloroplast (Cline et al. 1985). Binding represents a physiologically relevant step of the import, as pre-protein bound in absence of ATP can be chased into imported and processed mature protein when ATP is added. Using different ionophores and nucleotide triphosphate concentration, the energetic requirements of the protein import process was investigated in more detail. As just mentioned, pre-protein can reversely bind to the chloroplast membrane in the absence of ATP (binding stage, fig. 3, a). The second step is the irreversible binding of the pre-protein to the chloroplast envelope in the presence of 100 µM ATP. The pre-proteins arrested at this step were later called early import intermediates (fig. 3, b). Finally, complete import and processing of the transit peptide require 1-5 mM ATP (fig. 3 c). It was first thought that ATP was necessary outside the chloroplast, but experiments with an "ATP trap" (apyrase), showed that only stromal ATP is needed for the import. A role for a Proton Motive Force driving the import of protein was also excluded (Keegstra et al. 1989).

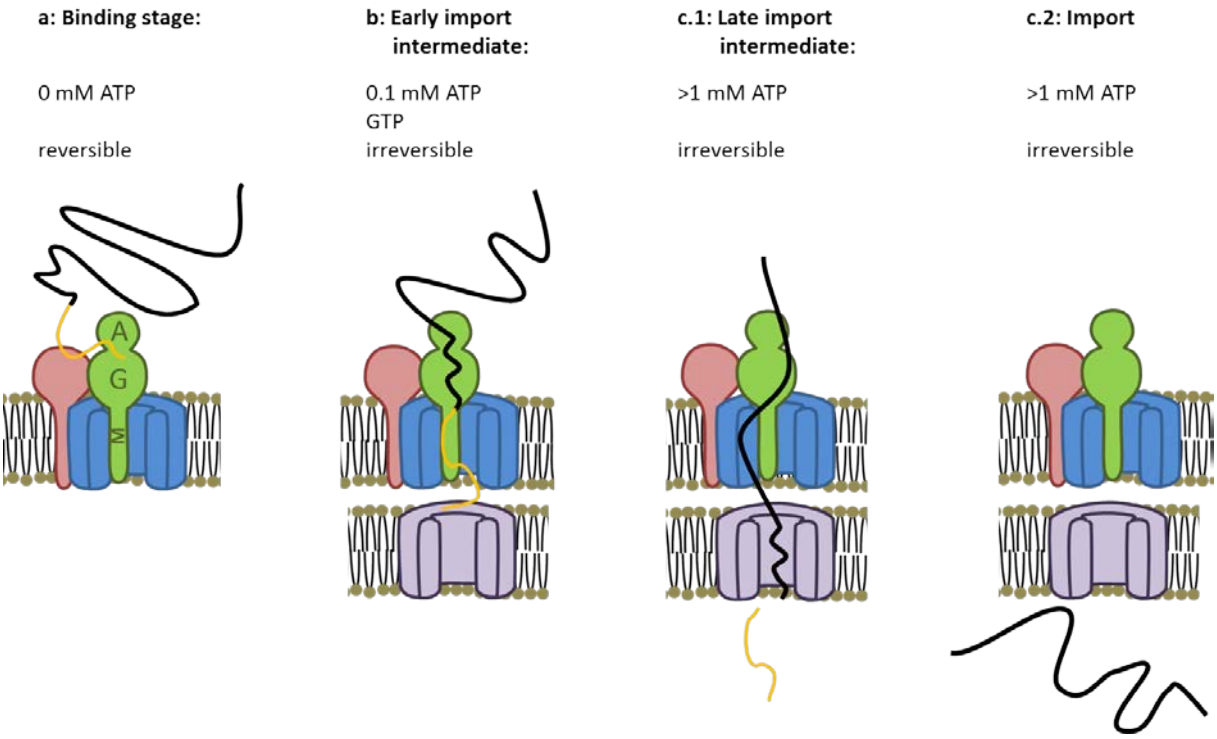


Figure 3: Energetic requirement of the import process: a) Binding stage: The pre-protein (black and orange line) first bind to Toc159 (green) and Toc34 (red) in an ATP independent way. **b)** Early import intermediate: in presence of 100 µM ATP, the pre-protein is committed in the Toc75 channel and makes contact with Tic proteins. This step depends on GTP hydrolysis. **c)** 1 mM ATP allows the full translocation of the pre-protein. Pre-proteins inserted into the import machinery but with the transit peptide (orange line) cleaved can be isolated and are called late import intermediates (**c.1**). Finally the pre-protein is fully imported in the stroma (**c.2**).

A general import pathway?

The precursor of the small subunit of rubisco (pSSU) was and is one of the most frequently used model precursors in import assays *in vitro*. This raised the question of whether other

import routes to the chloroplast than that used by pSSU exist. Several other precursors with different sub-chloroplastic localization were successfully imported under the same conditions as pSSU. Interestingly their import can be competitively inhibited by pSSU, suggesting that these proteins share the same import pathway (Perry et al. 1991; Schnell et al. 1991; Row 2001).

Identification of the putative receptors:

Protein factors at the chloroplast surface are likely involved in pre-protein binding, as protease (thermolysin) digestion of proteins exposed on the cytosolic side of the outer membrane abolish the ability of the chloroplast to bind a precursor protein (Cline et al. 1985; Friedman & Keegstra 1989). Different strategies were used to attempt the identification of the predicted outer membrane receptor for transit peptide binding. One of these consisted in raising an antibody against the pSSU transit peptide, assuming that the recognition site of this antibody will have steric properties similar to that of the putative transit peptide receptor. Another antibody is raised against this “mold” of the receptor and use to try to capture the real one (anti-idiotipic antibody) (Pain et al. 1990). Waegemann & Soll (1991) showed that some of the already identified OEP (outer envelope protein) co-migrated with a radiolabeled precursor on a sucrose density gradient. Another group managed to “tag” candidate proteins for the translocation apparatus by a label transfer experiment. They used a recombinant pSSU carrying a radiolabeled photoactivatable crosslinker. Once the pSSU engaged in the import process, the crosslinker was activated by UV light illumination, allowing the radiolabeled crosslinker to react with nearby proteins (Perry & Keegstra 1994). In 1993, Schnell & Blobel (1993) highlighted the existence a late import intermediates. Precursors were first stably bound at “low” concentrations of ATP (50 um) to form an early import intermediate. When the pre-proteins at the early import intermediate stage were chased into the stroma by “high” concentrations of ATP (5 mM), it was possible to obtain a late intermediate by stopping the reaction on ice (fig. 4, c). The precursor at the late import intermediate stage is already processed but still bound to the outer/inner membrane fraction of the chloroplast. By electron microscopy, the early and late import intermediates were shown to locate at so called “ contact sites”, where the inner and outer membrane of the envelope are touching (Schnell & Blobel 1993). Using a ProteinA-tagged SSU as pre-protein, Schnell et al. (1994) and Kessler et al. (1994) pulled-down the proteins bound to the early import intermediate and to the late import intermediate, from isolated envelopes using mild detergent solubilization. 4 proteins, IAP86, IAP34, IAP75, and IAP70 were associated with the early import intermediate (IAP: Import Associated Protein). IAP100 and IAP36 were co-purified with the late import intermediate, in addition to the 4 already found associated with the early intermediate. Some of these proteins had already been identified as potential components of the import machinery. However this experiment showed that the import intermediate forms a stable complex with the putative translocon proteins without crosslinking and even after the solubilization of the envelopes. In addition appropriate controls were carried out to exclude that these proteins are contaminations from the envelope, as may be argued for the other strategies (Kessler et al. 1994; Schnell et al. 1994). Moreover, starting from Edman sequenced peptides of IAP86, IAP34 and IAP75 it was possible to isolate cDNAs and derive the coding sequences of the three proteins. IAP86, IAP34 and IAP75 were shown to be associated in a complex, independently of the binding

with a pre-protein (Ma et al. 1996). In 1997, as the same import proteins were referred to with different names in the literature, a unified nomenclature was proposed. The outer and inner import proteins were named with the suffix Toc or Tic respectively (Translocon at the Outer/Inner envelope of the chloroplast) followed by their molecular mass. IAP86, IAP34, IAP75 and IAP110 became Toc86, Toc34, Toc75 and Tic110 respectively (fig. 4) (Schnell et al. 1997).

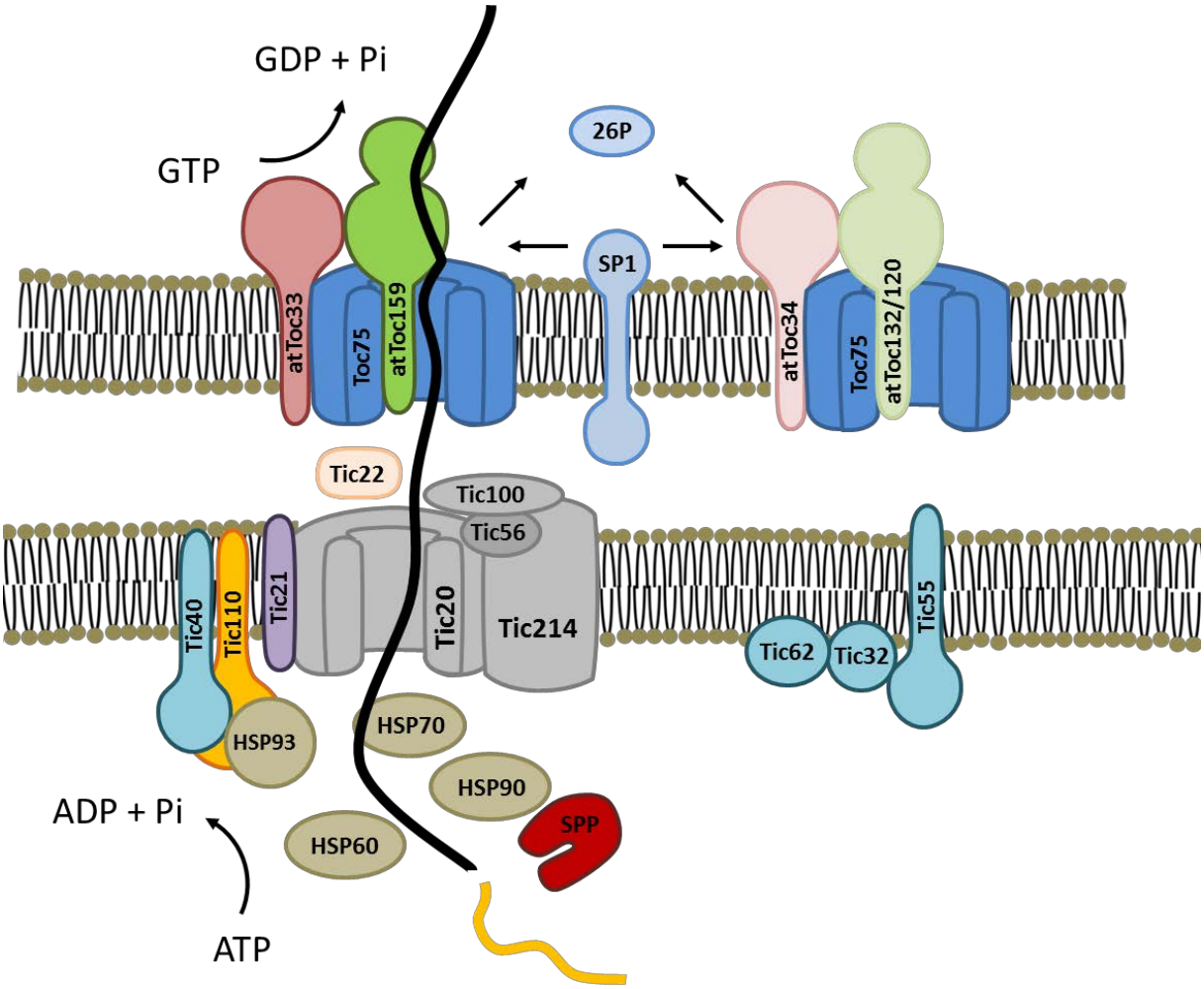


Figure 4 : Organization of the translocons at the outer and inner membrane of the chloroplast (Toc/Tic complexes). The Toc complex is composed of two receptors, Toc159 and Toc33 (psToc34) and a channel, Toc75 and preferentially imports photosynthetic proteins. Homologues of Toc159 and Toc33, Toc132/120 and Toc34 respectively, form a distinct complex which preferentially imports housekeeping proteins. At the inner membrane of the chloroplast, the pre-protein is imported through the Tic20 channel in association with either Tic40 and Tic110 or with Tic214, Tic100 and Tic56 (1 MD Tic complex, grey). Stromal chaperones (HSP70 and HSP90) are thought to assist the folding of the pre-protein and to provide the driving force for import, pulling the pre-protein in the stroma in an ATP-dependent way. Tic62, Tic32 and Tic55 might regulate the import depending on the redox potential of the stroma. The Stromal Processing Peptidase (SPP) cleaves the transit peptide. The E3 ubiquitin ligase SP1 mediates the targeted degradation of Toc proteins.

Sequence of the pre-protein interactions with the Toc/Tic proteins:

Label transfer crosslinking done with a pre-protein at the import intermediate stages described above provided information on the sequence of interactions of the pre-protein with the different component. During the energy-independent binding stage, the pre-protein interacted with Toc34 and Toc86, suggesting that these two proteins act as the pre-protein

receptors of the outer membrane translocon. The early import intermediate transit peptide interacted principally with Toc86 and Toc75, and The late import intermediate was mainly associated with Tic20 and Tic22(see below) (fig. 4) (Kouranov et al. 1998; Kouranov & Schnell 1997).

Biochemical vs genetic studies:

The mechanisms and components of the import apparatus were first studied on pea chloroplasts by biochemical methods. Along with the complete sequencing of *A. thaliana* genome and the arrival of reverse genetic techniques, the role of the different Toc/Tic components and their mechanisms were tested in vivo by analyzing the phenotypes of mutant lines and when possible the import properties of isolated mutant plastids. Many discoveries gained by biochemical means were confirmed, but the importance of some of the identified Toc/Tic protein and regulatory mechanisms were challenged by the complementary *in vivo* data. The next part of the introduction describes the various Toc/Tic components in more detail. This includes the components already mentioned above, as well as other component which were identified later.

Toc/Tic components:

Toc86-Toc159:

Toc159 was originally identified in peas as a smaller protein called Toc86 but the identification of the gene coding for the homologue of Toc86 in *A. thaliana* revealed that the pea Toc86 lacks a huge N-terminal portion and that the actual molecular weight of the full length protein was of 159 kDa (Bölter et al. 1998). Toc159 is inserted into the outer membrane protein by a 52 kDa C-terminal domain, the rest of the protein being exposed to the cytosol (fig.3-4). Its insertion in the membrane is GTP dependent. (Kessler et al. 1994; Hiltbrunner et al. 2001; Smith et al. 2002). The central part of Toc159 contains a conserved GTP-binding protein motif homologues to the one of Toc34. Bauer et al. (2000) identified and characterized the T-DNA insertion mutant of *toc159* in *A. thaliana*. The mutant was called *ppi2*, for plant plastid mutant n°2 as it has an albino phenotype and was seedling lethal on soil (fig.5, a). In *ppi2* chloroplasts fail to develop and have barely any thylakoid membranes and lack starch granules, indicating that they are non-functional (fig. 5, b). Essential photosynthetic protein like CAB and SSU are accumulated only at very low levels (fig, 5.c). This phenotype is consistent with a defect in import capacity but the genes encoding these proteins were also down-regulated. However, some photosynthetic as well as non-photosynthetic genes were normally transcribed and their protein products proteins were normally accumulated in *ppi2* plastids. Based on its sequence Toc159 appeared to have 3 domains, the A (acidic)-domain at the N-terminus highly enriched in acidic amino acid residue, the central G- (GTPase) domain, and the C-terminal M- (membrane) domain, anchoring the protein in the membrane (fig. 3, a). The M-domain is rather hydrophilic but it corresponds to the 52 kDa C-terminal fragment resistant to thermolysin digestion.

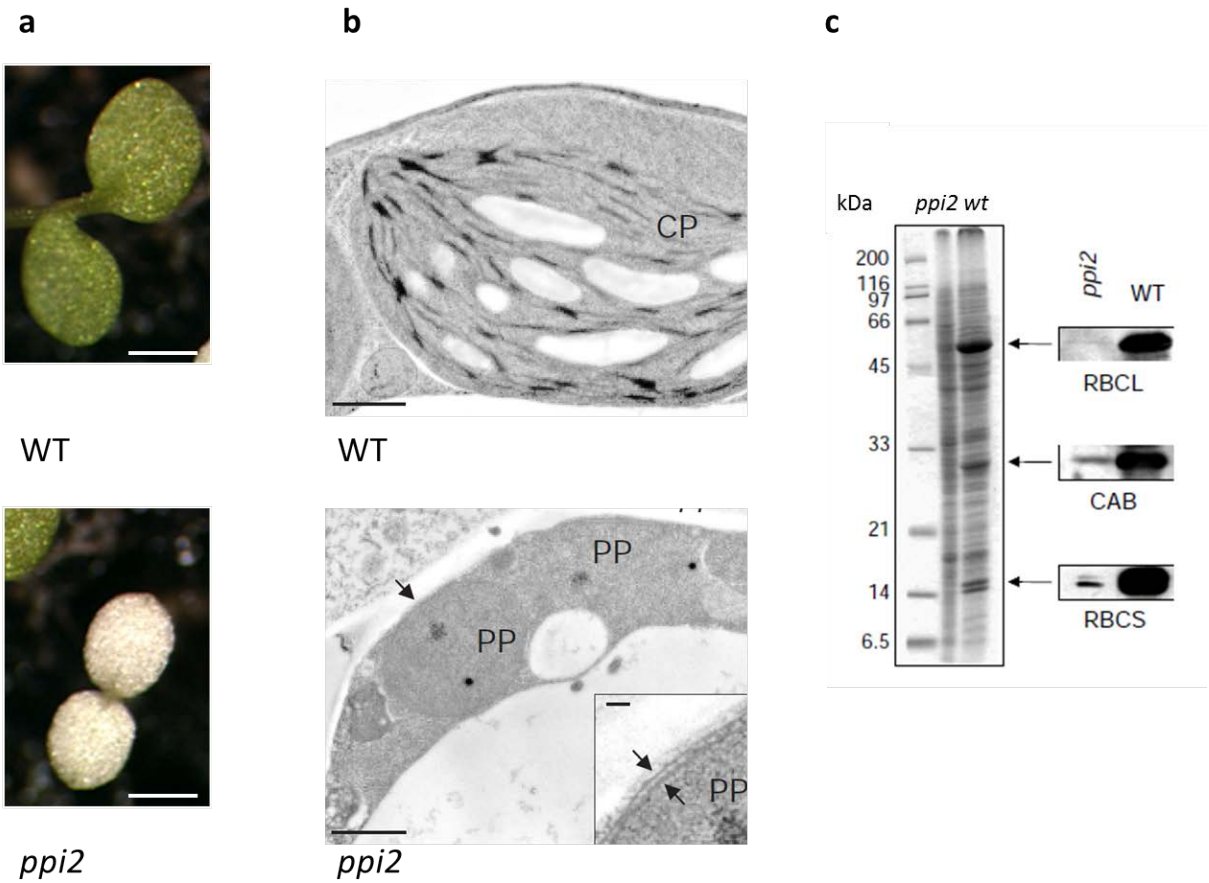


Figure 5: The *toc159* mutant has an albino phenotype. a) The *A. thaliana* mutant of Toc159 (*ppi2*, *plastid import mutant 2*) , one of the three major Toc components biochemically identified, has an albino phenotype, and is seedling lethal when grown on soil (**b**). In *ppi2,* plastids remain undifferentiated pro-plastids (PP), with no visible thylakoid membranes compared to the chloroplast (CP) of a wild-type plant. Scale bar, 0.5 µm. Inset: the double membrane of the pro-plastid. Scale bar, 0.05 µm (**c**). The *ppi2* mutant exhibits a strongly reduced accumulation of nucleus encoded photosynthetic proteins such as the small subunit of rubisco (RBCS), the chlorophyll a binding protein (CAB) but also of the chloroplast encoded large subunit of Rubisco (RbcL) (Bauer et al. 2002).

Three homologues of Toc159 were identified in the *A. thaliana* genome, Toc132, Toc120 and Toc90. Their G- and M-domains are highly conserved, whereas their A-domains strongly diverge in sequence and length (Bauer et al. 2000; Hiltbrunner et al. 2004). A structural study of the acidic domain of Toc159 demonstrated that it has the characteristic features of an intrinsically disordered protein (Richardson et al. 2009; Hernández Torres et al. 2007). Toc90 showed convergent properties with Toc159. Indeed, whereas the homozygous *toc90* mutant had no visible phenotype, the residual accumulation of photosynthetic proteins in Toc159 mutant was almost completely abolished in the *toc159/toc90* double mutant. Also, the *toc159/toc90* phenotype was more severe than that of the *toc159* mutant, but the accumulation of non-photosynthetic proteins was not affected. Moreover, overexpression of Toc90 in the *ppi2* background partially complemented the albino phenotype (Hiltbrunner et al. 2004; Infanger et al. 2011).

Toc34/Toc33:

Toc34 was identified by the same way as Toc159, in association with early and late ProteinA-tagged import intermediates (Kessler et al. 1994). Apart from a predicted transmembrane helix, Toc34 consists uniquely of a GTP binding domain homologous to that of Toc159. Toc34 inserts spontaneously in the outer envelope without the need for protein factors and it behaves as an integral membrane protein. Toc34 without its transmembrane domain couldn't insert in the outer membrane. Toc34 faces the cytosolic side of the outer envelope as it is completely sensitive to thermolysin (fig. 3-4). Toc34 purified from a heterologous system has a GTPase activity and native Toc34 can bind GTP in vivo (Kessler et al. 1994; Seedorf et al. 1995; Chen & Schnell 1997; Li & Chen 1997). The GTPase function of Toc34 is necessary for a proper insertion in the membrane (Chen & Schnell 1997). In absence of ATP, i.e. during the energy independent binding phase of import, Toc34 binds to the transit peptide of the pre-protein and this binding is suppressed in the presence of GTP (Kouranov & Schnell 1997). Contradictory results reported that binding of GTP by Toc34 strongly enhances its interaction with a transit peptide. In turn, the phosphorylation of Toc34 was proposed to inhibit the binding of GTP, consequently inhibiting the binding of the transit peptide (Sveshnikova, Soll, et al. 2000). A mutant of Toc33, the Arabidopsis orthologue of the pea Toc34, was discovered in a genetic screen for pale green mutants. It accumulates less chlorophyll than the wild type during its development and shows a reduced capacity for *in vitro* import of some proteins (Jarvis et al. 1998; Kubis et al. 2003) but this phenotype is much less strong than *ppi2*, the Toc159 mutant. *ppi3*, the mutant of Toc34, a second orthologue of pea Toc34 in *A. thaliana*, has no visible phenotype (Constan et al. 2004).

Based on complementation of the *ppi1* mutant controversial results were reported concerning the importance of the phosphorylation of Toc34 *in vivo*. Jarvis et al. (2006) transformed the *ppi1* mutant with Toc33 with two different mutations at the presumed phosphorylation site, one abolishing and the other mimicking phosphorylation. Both constructs complemented *ppi1* completely (Aronsson et al. 2006). However, Oreb et al. (2008) demonstrated that in young seedlings (5 days old) of the same lines, photosynthesis and chloroplast development was reduced (fig. 4).

Toc75:

Toc75 is embedded in the outer-membrane of the envelope although its amino acid sequence is largely hydrophilic and contains no predicted transmembrane helices (Schnell et al. 1994). The targeting of Toc75 to the outer membrane is unconventional. It has a bipartite transit peptide and it is first imported in the stroma where it is processed to an intermediate size. In a second step, it is directed to the outer membrane where it is processed to the mature size (Tranel et al. 1995; Tranel & Keegstra 1996; Inoue et al. 2001). From the beginning it was proposed to be the channel of the Toc complex. Several lines of evidence support this hypothesis, it interacts directly with the preproteins and electrophysiological studies showed it has the properties of a cation selective pore, reminiscent of those of beta-barrel channel proteins (Hinnah et al. 1997). In agreement with the absence of predicted transmembrane helices, structural homology modeling indicates a beta-barrel structure for Toc75, with 16 beta-sheets spanning the membrane (Sveshnikova, Grimm, et al. 2000; Hinnah et al. 2002). Two homologues of Toc75 were identified in *A. thaliana*, Toc75-III and

Toc75-IV. Toc75-III is strongly expressed during the young seedling stage, whereas Toc75-IV is more constantly expressed but at a lower level. The Toc75-III homozygous mutant is embryo lethal and Toc75-IV has no visible phenotype. These data support a major role for Toc75-III in the *A. thaliana* Toc complex (fig.4) (Baldwin et al. 2005).

Tic110:

Tic110 was found associated with a late import intermediate (Kessler et al. 1994). It is an integral membrane protein with two predicted transmembrane helixes near the N-terminus and the large C-terminal portion is soluble and exposed to the stroma (Kessler & Blobel 1996; Jackson et al. 1998; Inaba et al. 2003). Using small angle x-ray scattering and x-ray crystallography, Tsai et al. showed that Tic110 has typical HEAT repeats, known to serve as a protein scaffold in protein-protein interaction (Tsai et al. 2013; Hernández-Torres et al. 2014) and Tic110 has been shown to interact with the protein chaperone Cpn60 at the stromal side of the inner membrane in an ATP-dependent fashion (Kessler & Blobel 1996). Upon interaction with a pre-protein, Tic110 recruits Tic40 which in turn bind to Hsp93 (Chou et al. 2006). The Tic110 homozygous T-DNA mutant in *A. thaliana* is non-viable and embryo lethal. The knock-down of tic110 accumulated many different chloroplastic proteins to a lower degree than the wild type (Inaba et al. 2005). Moreover, a tic110 dominant negative mutant disturbs the assembly of the native Tic110 with Tic40 and Hsp93 (Inaba et al. 2005). Tic110 was proposed to act as a protein scaffold which recruits stromal chaperones upon pre-protein arrival in the stroma. An alternative role for Tic110 as the protein import channel at the inner membrane of chloroplast was proposed. Indeed, electrophysiological experiment revealed the pore like activity of Tic110 (Heins et al. 2002). In disagreement with the pore hypothesis, Tic110 can be expressed as a soluble protein when the two N-terminal transmembrane helixes are removed (fig.4) (Inaba et al. 2003)

Tic40:

Tic40 is anchored at the inner membrane of the envelope by a single membrane domain, with a large hydrophilic portion exposed to the stroma (Tripp et al. 2007). Tic40 correspond to the aforementioned pea IAP36, identified by co-purification with a recombinant pre-protein (Schnell et al. 1994). Tic40 was shown to co-precipitate with Toc75, Tic110 and Hsp93 and to be implicated in the same stage of the protein import process. This is supported by the fact that Tic40 and Hsp93 double mutant showed no additive effect compared to the single mutants (Stahl et al. 1999; Kovacheva et al. 2005). *A. thaliana* Tic40 null mutant has a pale phenotype and is impaired at the translocation step of protein import but the formation of the early import intermediate is not affected. The interaction with Tic110 is stimulated by pre-protein binding to Tic110 (Chou et al. 2003; Stahl et al. 1999). Together with Tic110, Tic40 is thought to be involved in the recruitment of stromal chaperone proteins to drive protein import or assist pre-protein folding upon arrival in the stroma. Tic40 is also implicated in the insertion from the stroma of newly imported inner membrane proteins (fig. 4) (Chiu & Li 2008).

Tic20 and Tic22:

Tic22 and Tic20 were discovered in pea by crosslinking to synthetic pre-protein early import or late import intermediate. Tic20 is an integral protein of the inner envelope and Tic22 is peripherally associated with the intermembrane side of the inner envelope. Tic22 and Tic20 interact in vivo (Kouranov & Schnell 1997; Kouranov et al. 1998). Tic20 has for homologues in *A. thaliana*, Tic20-i, Tic20-ii, Tic20-iv and Tic20-v. Tic20-i is the best studied homologue, it has an albino and seedling lethal phenotype and was proven to be essential four protein import (Chen 2002; Kasmati et al. 2011; Hirabayashi et al. 2011). It was proposed as a channel for the protein translocation at the inner membrane as it has a cation selective channel activity (Kovacs-Bogdan et al. 2011). Tic22 has two homologues in *A. thaliana*, Tic22-iii and Tic22-iv. Tic22-iii mutant has a light chlorotic phenotype while Tic22-iv mutant has no visible phenotype. Double Tic22-iii/Tic22iv mutant shows a more severe chlorotic phenotype than Tic22-iii mutant and chloroplast import assays revealed its reduced import competency suggesting a redundant function of both homologues (Rudolf et al. 2013). Tic22 was proposed to have a chaperon function (fig. 4) (Glaser et al. 2012).

Tic21:

Tic21 was isolated in A. thaliana as a result of a screen for import mutant. Tic21 is an integral protein at the inner membrane of the envelope. It has an albino phenotype and shows a reduced import rate. It is thought to be functionally similar to Tic20-1 (fig. 4) (Teng et al. 2006).

Chaperones and molecular motor:

Hsp70 was proposed to act as a chaperone in the cytosol to prevent aggregation or mis-folding of pre-proteins on their way to the chloroplast. Indeed, most transit peptides have an Hsp70 binding motif, and physical and direct interaction between transit peptide and Hsp70 was reported. In association with a 14-3-3 protein, Hsp70 was shown to bind specifically with the phosphorylated transit peptide, serving as a guidance complex for the delivery to the import apparatus (May 2000). However, this guidance complex seems to be dispensable *in vivo*. Indeed, a mutation of the transit peptide at the 14-3-3 binding site has no effect on the import of a protein (Rial et al. 2003). In pea, an Hsp70 was found to interact with the pre-protein in the intermembrane space. However, the *A. thaliana* homologues of this protein are not localized in the intermembrane space (Schnell et al. 1994; Ratnayake et al. 2008). Three stromal proteins were reported to interact with the Tic proteins or a late import intermediate: Hsp93 (or ClpC), Hsp70 and Cpn60. Hsp93 was found cross-linked to an early import intermediate and it was also co-precipitated with Toc complex components (Nielsen et al. 1997; Kouranov et al. 1998). Hsp93 is also known as Clp (Caseinolytic Protease) C and has a AAA+ (ATPase associated with various cellular activities) domain. In bacteria, ClpC forms hexameric rings with a central pore that can associate with two heptameric ClpP rings. The ClpC ring mediates the recognition and unfolding of a substrate proteins by threading it through the pore where it is handed to the ClpP ring for proteolysis. The threading is dependent on the ATPase activity. As it was also shown to interact with a transit peptide *in vitro* (Rosano et al. 2011), it was hypothesized that Hsp93 might act as molecular motor

which pulls pre-proteins in the stroma by threading it through an import channel in a ATP dependent way (Flores-Pérez & Jarvis 2012). Hsp93-iii and Hsp93-v were identified in *A. thaliana*. Hsp93-V has a pale phenotype whereas Hsp93-III has a wild type phenotype (Wardle et al. 2007). The current model proposes that upon arrival in the stroma, a pre-protein first binds to Tic110, Tic110 then recruits Tic40, and finally Tic40 recruits Hsp93, which pulls the protein in the stroma (Paila et al. 2015). Hsp70 helps folding and prevents aggregation of proteins and it binds to its substrate in an ATP-dependent way (Fink 1999). The *A. thaliana* null mutant of stromal Hsp70-1 and Hsp70-2 showed reduced import capacity. Hsp70 was co-precipitated together with Hsp93 and Tic110 with a late import intermediate (Chen & Li 2007). The *tic40/hsp70* double null mutant is seedling lethal suggesting that they might be involved in parallel functions during protein import (Su & Li 2010). In analogy to the role of Hsp70 in mitochondria, it was proposed that Hsp70 might be the motor pulling the pre-protein in the stroma, together with Hsp93. ATP hydrolysis by Hsp70 might induce a conformational change triggering the pulling of the bound pre-protein. An alternative would be a ratchet mechanism: thermal agitation drives the arrival of the pre-protein and the binding of successive Hsp70 chaperones prevent slipping back into the cytosol. Cpn60 are homologues of the GroEL bacterial chaperone (Bertsch et al. 1992). It forms a large complex composed of two Cpn60 heptameric ring with a Cpn10 heptameric cap. The client protein is folded in the cavity defined by the rings (Flores-Pérez & Jarvis 2012). Cpn60 interacts with Tic110 in an ATP-dependent way and it was co-precipitated with a recombinant late import intermediate (Tsugeki & Nishimura 1993; Kessler & Blobel 1996). Cpn60 is proposed to help the folding of newly imported proteins (fig. 4). Another chaperone, Hsp90 was co-purified with an *in vitro* imported pre-protein and the corresponding *A. thaliana* mutant is embryo-lethal. Hsp90 is thought to be a part of the chaperon pool mediating the import at the stromal side of the inner membrane (Inoue et al. 2013).

Stromal processing peptidase:

Stromal processing peptidase (SPP), a metalloprotease responsible for the cleavage of the transit peptide in the stroma was isolated and cloned in pea. It recognizes the 12 C-terminal amino acids of the transit peptide. This enzyme has a second protease activity, which further degrades the transit peptide after the removal from the mature protein (Richter & Lamppa 1999; Richter & Lamppa 2002). The removal of the transit peptide seems to be an essential step for the appropriate function of imported proteins, as *A. thaliana* homozygous mutant for SPP is embryo lethal and its down-regulation in pea leads to a chlorotic phenotype (fig. 4) (Zhong et al. 2003).

Toc/Tic complexes:

Toc86, Toc34 and Toc75 were co-purified bound to the same pre-protein but co-purification experiment with antibody showed that they interact independently of the presence of a pre-protein (Ma et al. 1996). A 880 kDa complex containing these three proteins was identified by BN-PAGE (Blue Native Polyacrylamide Gel Electrophoresis) with a proposed stoichiometry of 1:3:3 (Toc159:Toc34:Toc75) (Kikuchi et al. 2006). This trimeric complex is referred to as the Toc core complex. The analysis of proteins bound to a pre-protein by sucrose gradient

separation and BN-PAGE revealed the existence of 3 distinct complexes during the import. First, the pre-protein binds to a c.a. 800 kDa complex containing Toc159, Toc75 and Toc34. Later it bind a 1300 kDa complex containing Toc159, Toc34, Toc75 and proteins involved in import at the inner membrane or the stroma, Hsp70, Tic110 and Hsp93. Finally a complex bigger than 1300 kDa with an unknown protein composition associates with the pre-protein at the late stage of the import (Chen & Li 2007). Co-immunoprecipitation experiment showed that a fraction of Tic110, Tic40, Tic22, Tic20, Cpn60 and Hsp93 co-purify with the Toc core complex proteins even in the absence of a synthetic pre-protein (Nielsen et al. 1997; Kouranov et al. 1998). This suggests the existence of a supercomplex containing import proteins from the outer and inner membranes and the stroma. This supercomplex might coincide with the outer and inner membrane contact sites highlighted by electron microscopy (fig. 4) (Inoue et al. 2013).

The Tic 1 MD complex :

In 2009, Kikuchi and colleagues identified by BN-PAGE a 1 MD complex which co-migrated with a radiolabeled pre-protein. This complex was shown to contain Tic20 and in a lesser extent Tic21, but Tic110 was not present. The characterization of the complex led to the discovery of 3 yet unknown proteins, Tic214, Tic100 and Tic56, in addition to Tic20. Tic100, Tic56 and Tic20 *A. thaliana* null mutants all have an albino phenotype. Tic214 which is encoded by the chloroplastic gene ycf1 was shown to be essential for survival in *C. rheinardtii* and *N. tabacum* as homoplastomic lines for ycf1 mutation could not be isolated (Boudreau et al. 1997; Drescher et al. 2000). Tic20 was already proposed as a Tic channel and when reconstituted in lipid bilayer, the Tic 1MD complex shows a channel activity. The Toc core complex proteins and the 1 MD complex components were co-purified with a recombinant import intermediate pre-protein. The author proposed the 1MD Tic complex as the general Tic translocon for protein import at the inner envelope (fig. 4) (Kikuchi et al. 2009; S Kikuchi et al. 2013).

Role of the Toc159 and Toc33 GTPase in protein import:

Small GTPase:

Small GTPases constitute a wide family of proteins, homologous to the α subunit of the heterotrimeric G-protein. They can switch from an active GTP-bound state and an inactive GDP bound state by hydrolysis of GTP to GDP. The exchange of the resulting GDP for a new molecule of GTP reactivates the protein. The GTPase switch is usually controlled by two types of proteins, the GAPs (GTPase activating protein) and the GEFs (GDP exchange factor). Both GTP and GDP bound states are of equal importance for the function to be achieved. For instance, in nuclear transport, Ran GTPase in association with an import receptor protein directionally transports cargo protein between nucleus and cytosol. Ran in GTP bound state is responsible for the unloading of the cargo and it must return to the GDP bound state to allow the import receptor to load a new cargo (Cherfils & Zeghouf 2013).

Toc33 homodimer and Toc33/Toc159 heterodimer:

As mentioned previously, Toc159 and Toc34 have a predicted homologous GTP binding domain and bind GTP in vitro (Kessler et al. 1994). Both proteins belong to the TRAFAC (Translation factor related) GTPase class of the septin GTPase family (Leipe et al. 2002). Interestingly, a GTPase activity is necessary for the import, and more particularly for the formation of the early import intermediate (Olsen & Keegstra 1992; Kessler et al. 1994; Ma et al. 1996), as a non-hydrolysable analogue of GTP inhibits this step, but GTP hydrolysis is not required for the transition from the early intermediate to the translocation itself (Young et al. 1999). This initiated intensive investigation on the potential role of Toc159/Toc33 GTPase domain in the control of the import process. Sun et al. (2002), established the crystal structure of Toc34 in its GDP bound state, revealing that it exist as a homodimer. The GTP/GDP binding sites were shown to be at the interface of the dimer. Moreover, an arginine finger was in close proximity with the GDP of the binding partner, reminiscent of the mode of action of a GAP protein. It was thus proposed that the two member of the homodimer mutually act as GAP proteins, by the mean of the arginine finger (Sun et al. 2002). However, it appeared that the arginine finger has no role in activation of the GTPase activity of toc34, but it is needed for the dimerization (Weibel et al. 2003). The transit peptide of pre-protein was shown to increase the Toc34 GTP hydrolysis rate (Schleiff et al. 2003). It was later shown that binding of the transit peptide did not increase the GTPase activity, but caused the Toc34 homodimer to dissociate thus allowing the exchange of GDP to GTP, indirectly promoting the binding and the hydrolysis of a new GTP (Oreb et al. 2011). Toc33 was also shown to dimerize withToc159 through a homotypic interaction of their respective G-domain. The existence of Toc33 homodimers and toc33/toc159 heterodimers was shown in *vivo using* a split-ubiquitin system (Hiltbrunner 2001; Rahim et al. 2009). In opposition to the Toc33 homodimer, the binding of transit peptide reinforces the stability of theToc159/Toc33 interaction. Different but overlapping region of the transit peptide were shown to interact directly with Toc34 and Toc159 G-domain (Becker et al. 2004). The role of the dimerization of Toc33 was tested in *A. thaliana* by complementing *ppi1* (Toc33 mutant) with a dimerization Toc33 mutant unable to form homodimer or heterodimer with Toc159. The *ppi1* phenotype was fully complemented, but *in vitro* import experiments showed that the translocation step of import was impaired, the binding being unaffected. A similar experiment showed that mutated Toc33 unable to bind GTP complemented *ppi1*, but reduced the import competency to 50% of the wild type.

Effect of Toc159 GTPase mutation in vivo:

A version of Toc159 carrying three point mutations in the G1-motif was unable to hydrolyze GTP and did not complement *ppi2* (J Bauer et al. 2002). This may be due to an effect on GTPase properties on the one hand or to structural effects of the triple point mutations on the G-domain on the other. Using a variety of single point mutations in G-motifs, it was later discovered that Toc159 mutated for the GTP hydrolyzing and GTP-binding function can complement the ppi2 phenotype (Smith et al. 2002; Wang et al. 2008; Birgit Agne et al. 2009). Interestingly, the mutation inhibiting GTP hydrolysis increased pre-protein *in vitro* import efficiency, whereas the mutation abolishing GTP-binding reduced pre-protein import efficiency suggesting that Toc159 in its GTP bound state is necessary for import. More

particularly, it was shown that the enhanced import activity of the GTP hydrolysis Toc159 mutant is due to an increased affinity for the pre-protein at the binding stage, before the formation of the early import intermediate. The early intermediate formation in this mutant was still sensitive to non-hydrolysable GTP, suggesting a role for the Toc33 GTPase activity (Feller et al. 2008). Toc33 and Toc159 are not functionally redundant but it might be that their GTP binding and GTPase functions are, explaining the relatively limited effect of individual mutations. So far the effects of mutations in both Toc159 and Toc33 GTP binding or GTPase activity has not been investigated (Paila et al. 2015).

Model of the Toc/Tic protein import process (fig. 4):

A sequential model for pre-protein recognition, initiation of the translocation and translocation itself was established from the aforementioned data. I) Toc159 and Toc33 are both in their GDP bound state, and Toc33 is in homodimer form. ii) The transit peptide of the pre-protein interact with both Toc159 and Toc33, disrupting the Toc33 homodimer and allowing the exchange of GDP to GTP. The disruption of the Toc33 homodimer and the binding of transit peptide favor the heterodimerization of Toc33 with Toc159 which presumably also shift to GTP bound state. This GTP bound Toc33/Toc159 heterodimer represents an activate state ready to commit pre-protein in the Toc75 channel. iii) GTP hydrolysis triggers the commitment of the pre-protein to the intermembrane space and the Tic proteins. Molecular chaperones in the intermembrane space might facilitate the interaction with the Tic proteins. Once the pre-protein reaches the stroma, it is pulled through the inner membrane in an ATP dependent way. The translocation through the inner membrane remains unclear and different models were proposed based on the known interaction of Tic proteins with a pre-protein or between each other. Tic110, Tic20, Tic21 and Tic22 might form a complex with Tic20 as the channel and Tic110 recruiting together with Tic40 the stromal chaperons, the 1MDa Tic complex forming a second and distinct import unit. Alternatively, only one import unit including all the Tic components mediates the import with Tic20 as the channel. The 1 MDa complex isolated by Kikuchi et al. didn't co-precipitate with Tic110, but in another study Tic20 and Tic110 were both shown to co-precipitate with Toc34, suggesting at least an indirect interaction (Paila et al. 2015).

Evolution of the import apparatus:

Toc75, Tic20 and Tic21 but not Toc159, Tic110 and Toc34 were shown to share a common origin with proteins of *Synechocystis*, a cyanobacterium. This suggests that the Toc/Tic proteins have a dual origin. As SynToc75 is involved in protein secretion, it is postulated that it served as a basis for the primitive import apparatus, during the early time of the endosymbiosis (Reumann et al. 1999; Reumann & Keegstra 1999).

Transit peptide, a hidden specificity:

As mentioned above, the transit peptide of chloroplastic protein is sufficient and necessary for proper targeting. Several attempts to identify a consensus sequence for the specific

import into the chloroplast obtained limited success. The transit peptides vary much in length, from 20 to 100 amino acids and their only common feature seems to be the high abundance of hydroxylated residue and the tendency to be positively charged, as acidic amino acids are poorly represented (Keegstra et al. 1989).

In mitochondria, the targeting sequence (the equivalent of the transit peptide) of the pre-protein forms alpha-helices which are necessary for a proper targeting (Duby et al. 2003). This is not the case for the chloroplastic pre-protein, as transit peptides were shown to be unstructured in aqueous solvent (von Heijne & Nishikawa 1991; Wienk et al. 1999). However, in organic solvent or in contact with membranes the formation of alpha-helices is induced (Buck 1998; Krimm et al. 1999; Wienk et al. 2000). These secondary structures might be induced in contact of the chloroplast envelope and have role in the specificity of chloroplast targeting. Interestingly, different cases of dual targeting of a protein to both the chloroplast and the mitochondria were reported, suggesting that mitochondria and chloroplast targeting determinants are not exclusive (Carrie et al. 2009). The role of the different parts of the transit peptide was investigated by the in vitro import of pre-proteins with partial deletions in the transit peptide. It showed that the N-terminal part of the transit peptide is crucial for the import. Deletion in the middle part reduces the import rate whereas the C-terminal part seems to be needed for the proper cleavage of the transit peptide (Rensink 1998). More precisely it was shown that the first part and the middle part of the transit peptide are needed for the import at the outer membrane and the import at the inner membrane respectively, suggesting a modular organization of the transit peptide (Rensink et al. 2000). Recent work showed that different functional classes of transit peptides exist, with different import specificity. A model was proposed where the transit peptide is divided into modules, each responsible for the interaction with a specific Toc/Tic protein. The different combinations of these modules determine the specificity of the transit peptide and might explain the difficulty to isolate a canonical transit peptide sequence (Li & Teng 2013). The transit peptide sequences are thought to have emerged from cyanobacterial sequences of proteins secreted by Syn75, the cyanobacterial homologue of Toc75. It was proposed that the high variability of the transit peptide sequence might have emerged by the shuffling of exons of bacterial origin (Reumann & Keegstra 1999). Although no consensus sequence was defined for the chloroplast transit peptide, it is possible to predict the cleavage site hence the length of a transit peptide of a protein, by training a neural network with a set of known transit peptides (Emanuelsson et al. 1999).

Selectivity and regulation of the import activity:

Developmental and environmental variations of the import efficiency:

The selective import machinery of the chloroplast ensures that the appropriate set of proteins is imported and prevents proteins destined to other cellular compartments to enter. The chloroplast proteome is primarily shaped by transcriptional control. Indeed, during photomorphogenis, photosynthetic genes are strongly up-regulated and their protein product rapidly accumulate (Ma et al. 2001). However, there are evidences that the import rate is also varying through the development of the plant and in response to environmental changes. Indeed, Dahlin & Cline (1991) showed that the import competence of chloroplasts

is the highest in young developing tissues and declines thereafter. In etioplasts, import competence also declines through development but dramatically increases upon exposure to light (Dahlin & Cline 1991). Teng et al. (2012) revealed a more complex regulation of import through developmental stages of the plant. In peas, they identified three distinct import patterns. Some proteins are more efficiently imported in older leaves, other are less efficiently imported while the import rate of some proteins remains constant. The transit peptides account for this differential import rate, as transit peptide swapping changes the import pattern of a given protein. The relevance of this regulation was assessed in vivo. Indeed, Tic40 whose import rate increases in older leaves, cannot fully complement its own mutation when fused to a SSU transit peptide, a protein whose import rate decrease in old plants. Two consecutive positive charges necessary for the higher import in older leaves were identified (Teng et al. 2012).

Different plastid types have overlapping but different import specificity:

High and low growth temperatures were also shown to decrease import capacity of the chloroplast, correlated with the expression of some of the Toc/Tic proteins (Dutta et al. 2009). Different types of plastids show a differential selectivity toward client proteins. Indeed, 3 photosynthetic and 2 non-photosynthetic proteins were shown to be imported into chloroplasts, whereas only the non-photosynthetic proteins were imported in root plastid (Yan et al. 2006). Other examples of differential import selectivity were highlighted (Wan et al. 1995; Wan et al. 1996), like between pollen amyloplasts, endoderm plastids and chloroplasts (Primavesi et al. 2008; Li & Teng 2013).

Specialized import unit in *A. thaliana*:

The basis of the import specificity and the potential link with the different Toc/Tic homologues was investigated in *A. thaliana*. In *A. thaliana*, different homologues were identified for the pea Toc159, Toc75, Toc34, Tic20, Tic22, Hsp93 and Hsp70 (Jackson-Constan & Keegstra 2001). These homologues have differential anatomical and developmental expression patterns. As an example, Toc159 and Toc33 are more expressed in photosynthetic tissue whereas Toc120 and Toc132 have a more uniform expression. Interestingly, Toc159 and Toc33 mutants are affected in biogenesis of the chloroplast while Toc132 and Toc120 mutants have shorter roots (Bauer et al. 2000; Jarvis et al. 1998). This raised the hypothesis that the different Toc components might be specialized in the import of a subset of chloroplastic proteins. Genetic studies support the functional difference of Toc159 homologues. Toc132 over expression cannot complement the *ppi2 mutant*. Toc132 and Toc120 homozygous mutants have no visible phenotype butToc132/Toc120 double mutant is embryo lethal, suggesting a redundant function (Kubis et al. 2004). Moreover, co-precipitation experiment showed that Toc132 and Toc120 formed a complex with Toc34 while Toc159 associates preferentially with Toc33 (fig. 4) (Ivanova et al. 2004). *In vitro* binding experiments highlighted the differential affinity of Toc132 and Toc159 for different pre-proteins. Toc159 and Toc132 bind preferentially the precursor of the ferredoxin photosynthetic protein and the E1α non-photosynthetic protein respectively. The affinity is conditioned by the A-domain of the Toc159 homologue. Indeed, the removal of A-domain abolishes the preference of Toc159 and Toc132 for one type of protein and exchange of A-

domain between Toc132 and Toc159 results in the exchange of the specificity. Interestingly, Toc132 with a Toc159 A-domain can complement *ppi2* which is not the case for the wt Toc132 (Inoue et al. 2010). Using yeast two hybrid quantitative assay, Dutta et. al quantified the affinity of Toc159 and Toc132 with a broader set of photo-synthetic and non-photosynthetic proteins. Toc132 and Toc159 without A-domain had similar affinities for the different proteins, but Toc132 with A-domain had a stronger and weaker affinity for non-photosynthetic and photosynthetic proteins respectively (Dutta et al. 2014). The proteome analysis of the *ppi2* mutant revealed that the specificity of Toc159 is not exclusively directed toward photosynthetic proteins. Indeed, some photosynthetic protein are not or less imported in *ppi2* but others are normally or more imported than in wild type (Bischof et al. 2011). The picture emerging is that Toc159 and Toc132 have shared sets of client proteins but have differential import efficiency for subsets of proteins and A-domains of the different Toc159 homologues condition this import efficiency. The existence of different protein import apparatus reflects the different needs in different tissues, but it could also prevent less abundant non-photosynthetic proteins from being overwhelmed by the bulk of photosynthetic proteins when competing for import.

Post-translational regulation of the import:

The changes in import efficiency are likely due to the transcriptional control of the different Toc/Tic proteins, but some post-translational mechanisms were also shown to regulate import. Ling et al. (2012) showed that the composition of the import proteins is partially regulated by targeted degradation. SP1, an ubiquitin E3 ligase was discovered in a screen for suppressors of the *ppi1* phenotype. SP1 was shown to be implicated in the ubiquitination of different Toc components. In the SP1 mutant, the transition from etioplast to chloroplast was impaired, proving the importance of adapted protein import machinery in chloroplast differentiation (Ling et al. 2012). It was shown that cold decreases *in vitro* import efficiency, due to a decreased ATP availability (Leheny & Theg 1994). The concentration of GTP is less critical as GTPases have a high affinity for GTP (Bourne et al. 1990), but an external control of Toc159/Toc33 of GTPase by another factor than the pre-protein transit peptide might play a role in regulation of protein import. The redox potential also has an influence on the import efficiency *in vitro*. Two mechanisms were highlighted. First, the reduction of disulfide bridges between the Toc complex proteins enhances the import rate whereas oxidation blocks it. Secondly, an increase of the NADP+/NADPH ratio has a positive influence on the import rate. The sensing of the NADP+/NADPH status is thought to be mediated by Tic32 and Tic62, two proteins which contain NADP(H) binding site. These two proteins associate to other Tic components upon NADP+ binding. Tic55 is thought to be involved in redox regulation of the import (Pilon et al. 1992; Seedorf & Soll 1995; Stengel et al. 2008; Stengel et al. 2009; Kovács-Bogdán et al. 2010).

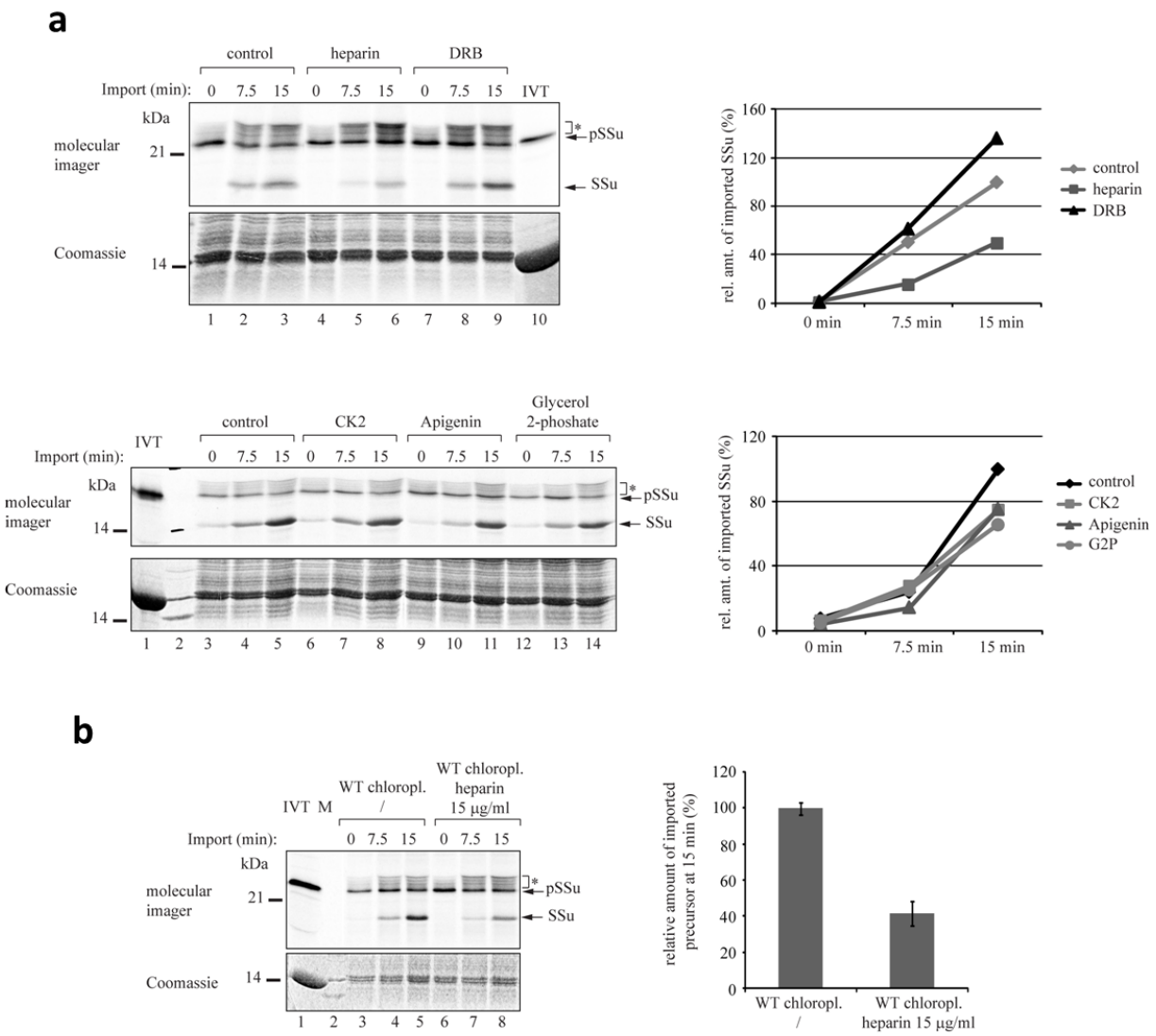


Figure 6 : Inhibition of in vitro chloroplast protein import by heparin. a) Isolated Arabidopsis chloroplasts were pre-incubated with or without 15 mg/ml heparin, 6 mM DRB, 3 units/ml recombinant maize CK2 α-subunit, 50 µM apigenin, or 10 mM glycerol-2-phosphate for 20 min at 25 C in the dark. Then, *in vitro* translated, [35S] Met-labeled preprotein of the small subunit of Rubisco (pSSU) was added and import was allowed to proceed for 0, 7.5, and 15 min. b) Quantification of the effect of heparin on chloroplast protein import. The graph shows the quantification of the amount of imported SSU at 15 min in three independent experiments. In both panels, the amount of SSU imported into wild-type (WT) chloroplasts without the addition of inhibitor at 15 min was set to 100%. IVT, in vitro translate; M, molecular mass standard. Asterisks indicate pSSU modified in the course of the import reactions.(Agne et al. 2010)

A role for A-domain post-translational modification?

Phosphorylation of the A-domain:

Many phosphorylation sites in the A-domain of Toc159 were identified by different mass spectrometry studies (de la Fuente van Bentem et al. 2008; Whiteman et al. 2008; Reiland et al. 2009; Agne et al. 2010). The *in vivo* phosphorylation status of the A-domain was also highlighted by ProQ diamond staining of a purified recombinant Toc159 expressed in *A. thaliana* treated with or without phosphatase. Moreover, a chloroplast membrane extract showed a kinase activity on A-domain heterologously expressed in bacteria. Different proteomic analyses of Toc159 identified 12 phosphorylated peptides in Toc159, exclusively in the A-domain. Most of the A-domain phosphorylation sites are predicted to be the target

of Casein Kinase 2 (CK2) class of kinase and CK2 phosphorylates *in vitro* recombinant A-domain expressed in bacteria. Taken together, these data suggests that there should be one or more kinases present in the chloroplast envelope which phosphorylates the A-domain of Toc159. Even though such an activity was never experimentally demonstrated, it is likely that A- domain is also the substrate of one or more phosphatase. Phosphorylation and dephosphorylation are a well-known switch mechanism which activate and deactivate a target protein. Heparin, a CK2 inhibitor was shown to reduce *in vitro* import competency of *A. thaliana* chloroplasts, but other kinase inhibitors (DRB and apigenin), a kinase (recombinant maize CK2) and a phosphatase inhibitor (glycerol-2-phosphate) had no effects (fig. 6, a-b). Heparin has a quite broad target spectrum and it could be that the import reduction is only an indirect effect. Nevertheless, it is tempting to conclude that import process might be regulated through phosphorylation events (Agne et al. 2010).

The Toc159 A-domain exists as a soluble fragment:

As mentioned above, Toc159 in pea was first called Toc86, as it shows a substantially smaller size than the one predicted from the DNA sequence. In *A. thaliana*, two bands specifically react with an anti-Toc159 antibody. The bigger band appears above 200 kDa and is present only in the membrane fraction of a plant extract and correspond to the full length Toc159.

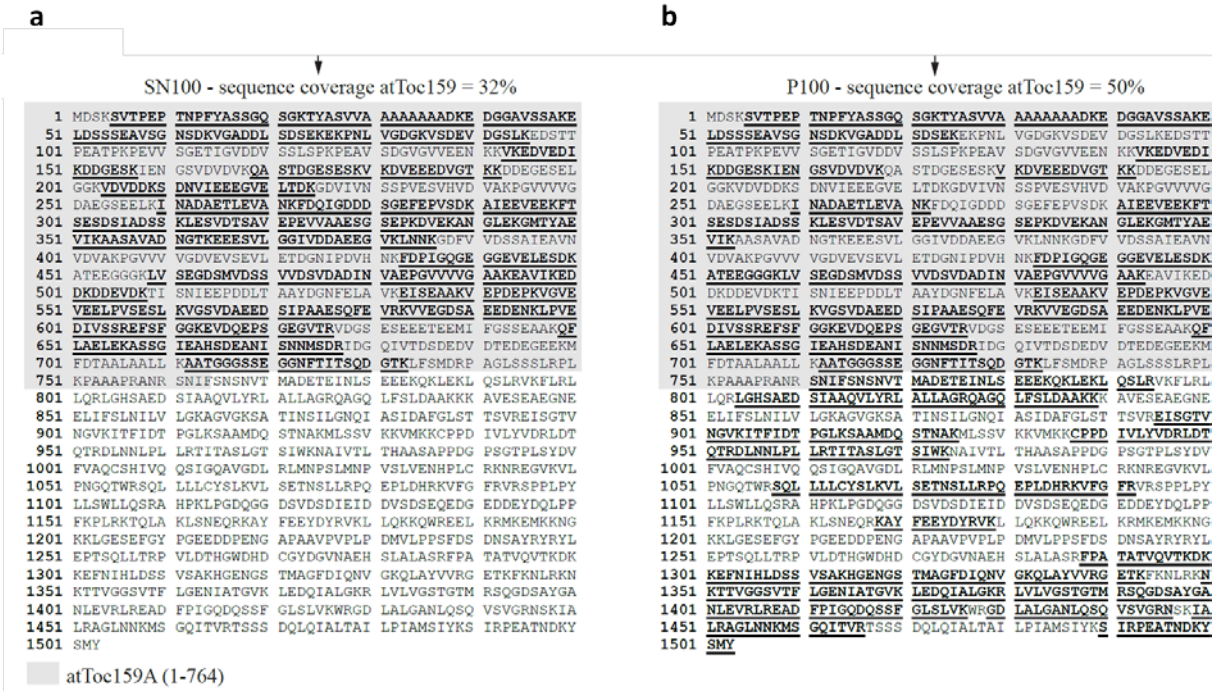


Figure 7: Toc159 A-domain exist as a soluble fragment: Recombinant N-terminally TAP-tagged Toc159 expressed in *A. thaliana* was affinity purified from soluble and membrane fractions of a plant extract. The eluates of the affinity purification were analyzed by mass-spectrometry. Sequences of the Toc159 peptides identified by mass-spectrometry in the eluate of the soluble fraction (a) or the membrane fraction (b) are shown. Amino acids of the peptides identified are in bold and underlined. Grey shading indicates the sequence corresponding to the A-domain.

The smaller band, between the 116 and 200 kDa marker bands, is found both in the soluble and membrane fraction of a plant extract (fig. 9, results section) and was first thought to

correspond to the GM-domain, with the A-domain being un-specifically degraded. The mass spectrometry analysis of the lower soluble band revealed that it corresponds to the Toc159 A-domain (fig.7, a-b). The existence of a cleaved but intact A-domain suggest a specific proteolytic event instead of unspecific degradation (Agne et al. 2010).

Aim of the thesis:

The chloroplast protein import process and its components were extensively studied and are rather well understood. After the identification of the main Toc/Tic components, the research extended to the comprehension of the regulation of the import and how the specificity is achieved in regard with changing need, depending on environmental and developmental constraints. The role of the differential expression of the different Toc/Tic homologues was highlighted. Post-translational events like targeted degradation and variation of the redox potential were shown to regulate the import apparatus, too. Interestingly, two post-translational modifications affecting the A-domain of Toc159 were highlighted. Although A-domain was shown to be dispensable for the complementation of Toc159 mutation (Agne et al. 2009), it is a key determinant of protein client specificity of Toc159 and its homologues. Post translational modifications of the A-domain might alter the affinity of Toc159 toward client pre-protein thus regulating protein import. The importance of the Toc159/Toc34 GTPase functions remains unclear. However, it can be postulated that the regulation of the Toc159/Toc33 GTPase activities through GEF or GAP might represent a regulation mechanism of protein import. The aim of this thesis was the identification and characterization of proteins which might interact and potentially regulate the Toc core complex. Particularly, potential proteases and kinases responsible for the cleavage or the phosphorylation of the Toc159 A-domain, or GTPase-related proteins were looked for. The proteins co-precipitating with a tagged Toc159 expressed *in A. thaliana* were analyzed by mass spectrometry. The list of candidates was sorted by abundance, predicted location and predicted functions. Three proteins were selected for further characterizations. This thesis reports the characterization of one of this candidates and its corresponding mutant (Emb2004/At1G10510). In addition, contributions were made for the localization and confirmation of the interaction with Toc159 of the two other candidates (AT4G32250 and AT5G01590/Tic56). Please note that when AT5G01590/Tic56 was selected as an interesting candidate, its presence in the 1MDa Tic complex was not yet reported (Shingo Kikuchi et al. 2013).

Material and methods :

Material:

Plants lines:

The A. thaliana (L.) Heynh. var. Columbia 2 (Col2) ecotype was used as the background of the Emb2004-TAP plant line.

The TAP-Toc159:*ppi2* and TAP:wt plants were in the Wassilewskija ecotype background and were already described (Agne, 2009).

Bacteria strains:

Cloning of plasmids was done in DH5α (Invitrogen AG, Basel). Protein over-expression was done in BL21 (DE3) (Novagen Inc, Madison WI, USA).The *Agrobacterium tumefaciens* strain C58 was used for the stable transformation of *A. thaliana* transient expression in *N.benthamiana.*

Yeast strains:

The AH109 and Y187 *S. cerevisiae* yeast strains were obtained from Clontech.

AH109: *MATα, trp1-901, leu2-3, 112, ura3-52, his3-20, gal4Δ, gal80Δ, LYS::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2, URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ, MEL1*

Y187: *MATα, ura3-52, his3-200, ade2-101, trp-901, leu2-3, 112, gal4Δ, met-, gal80 Δ, URA3::GAL1$_{UAS}$-GAL1TATA-lacZ, MEL1*

Chemicals:

All chemicals were purchased from Sigma-Aldrich or AppliChem when not specified.

Plasmids:

pET21d is a vector for bacterial expression of N- or C-terminally His tagged proteins with, a T7 promotor and an ampicillin resistance cassette (Novagen, EMB Millipore).

pEARLYGATE 205 (pEG205) is a plant expression vector base on the Gateway® technology for the expression of C-terminally TAP tagged fusion proteins, under the control of the CaMV 35S promotor.

pEARLYGATE 101 (pEG101) is a plant expression vector base on the Gateway® technology for the expression of C-terminally YFP tagged fusion proteins, under the control of the CaMV 35S promotor

pDONR™221 is a Gateway® "entry" vector enabling the cloning of a sequence of interest in an "acceptor" vector via recombination (life technologies Europe, Zug).

pGBKT7 and PGADT7 are vectors for the expression in yeast of N-terminal fusion proteins with the Gal4 binding and activating domains respectively. They have TRP1 and LEU2 selection genes, respectively.

pETM40 is a bacterial expression vector for the production of an N-terminal fusion with a Maltose Binding Protein (MBP) (EMBL).

Antibodies:

Anti-atToc15A (Bauer et al. (2000), anti-atToc75 (Hiltbrunner et al. (2001), anti-Toc33 (Agne et al. (2009), anti-atPGL35 (Vidi et al. 2006), anti-atTic20 have previously been described.

The following antibodies were generous gifts: anti-Tic20 (Dr Hsou-Min Li, Academia Sinica, Taiwan) (Teng et al. 2006), anti-atTic56 (Dr B. Agne, Martin-Luther-Universität Halle-Wittenberg, Germany)(Koehler et al., 2015), anti-CAB (Dr. K. Apel, ETH, Zürich) anti-SSU (Dr. Pia Stieger, Université de Neuchâtel) and anti-AT4G32250/KOC1 (Mónica Arias Maldonado, Université de Neuchâtel).

Anti-AT4G32250/KOC1 was produced and purified by Monica Arias Maldonado (Université de Neuchâtel). A recombinant His_6-tagged fragment of AT4G32250 was used rabbit immunisation (Eurogentech) and specific IgG was affinity purified against the same fragment covalently linked to Affigel-10 resin (BIO-RAD). By immunoblotting, the purified antibody detects one band in a wild type leaf extract and corresponds to the predicted size of At4G32250 (66 kDa). The band as not detected in an *at4g32250* null mutant leaf extract.

Rabbit IgG was purchased from MP Biomedical (Santa Ana, CA, USA, ref. 55944), anti-CBP from GenScript (Piscataway, NJ, USA; ref. A00635), anti-actin from Sigma-Aldrich (St. Louis, MO, USA, ref. A0480), anti-RBCS, anti-LHCB2 and anti-atTic40 from Agrisera (Vännäs, Sweden, ref. AS07259, AS01003 and AS10709), anti c-myc from Roche (Basel, ref.11667149001)

Anti-Emb2004 antibody:

The full length *Emb2004* DNA sequence was cloned in the bacterial expression vector pETM40 plasmid to obtain a C-terminal fusion with the Maltose Binding Protein (MBP). As the recombinant protein was not soluble, the inclusion bodies containing the proteins were purified using 4 M urea (*A. McGettrick and M. Worrall*, Protein purification Protocols: Second edition, Human Press Inc., 2004). The purified inclusion bodies were dried and sent to Eurogenetech for a rabbit immunization program. The immune serum was purified against urea solubilized MBP-Emb2004 covalently linked to an Affi-Gel15 column (BIO-RAD).

Methods:

Plant methods:

Plants media:

½ Murashige and Skoog medium (1/2MS) solid medium: 0.8% agarose for plants (Duchefa), 2.2 g of Murashige and Skoog salts with vitamins (MS, Duchefa), 1% sucrose, 2.5 mM MES KOH pH 5.7. Autoclaved at 121°C, 1 atm for 20 min.

Plants were sown on sandy soil ("Rasenerde Top Dressing," Ricoter AG) soaked with a 0.25% *B. thuringensis* extract (SolBac) and incubated under long day conditions (16-h light, 8-h dark, 120 $\mu mol*m^{-2}*s^{-1}$, 21 °C).

Growth on agar media:

200 µl of *A. thaliana* seeds were used for 10 14 cm diameter petri dishes. This yielded approx. 10 g of seedlings after 2-3 weeks of growth on a 1/2 MS medium. The seeds were placed in a 1.5 ml microfuge tube and shaken at 250 rpm for 5 min in 70% ethanol with 0.05% TX-100, and 10 min in 1 ml of 100% ethanol. The seeds were spread on sterile filter paper under a sterile hood and allowed to dry. The dried seeds were sprinkled on the agarose medium. The petri dishes were taped with surgical tape (Micropore, 3M) and stratified at 4°C for 48h. The plants were grown with an 8/16 h light/night cycle in a growth chamber with controlled light, temperature and relative humidity parameters (120 µmol*m^{-2}*s^{-1}, 21/18 °C and 55% RH).

50 µM of Glufosinate ammonium (Dr. Ehrenstorfer GmbH, Germany) was added after the sterilisation of the medium when selection for this herbicide was needed.

Growth on soil:

The seeds of *A. thaliana* were sprinkled on individual pots filled with sandy soil (Top dressing, Ricoter) soaked with 0.25% of *B. thurigensis* extract (SolBac). The seeds were stratified for 48h at 4°C and grown under the same conditions described for the plants grown on ½ MS agarose media.

Chloroplasts isolation:

Isolation of A. thaliana chloroplasts:

The protocol used for *A. thaliana* chloroplast isolation is an adaptation of the protocol published by Smith et al. (2003). Approx. 10 g of 2 weeks old *A. thaliana* seedlings grown on agarose medium were harvested with a scalpel blade, transferred to a 9 cm petri dish and washed in 20 ml of cold digestion buffer (400 mM sorbitol, 20 mM MES-KOH pH 5.2, 0.5 mM CaCl2). The seedlings were chopped for 1 min with a scalpel blade and drained from the buffer after 5 min on ice. An additional washing step was done and 20 ml of digestion buffer containing 3% cellulase and 0.6% macerozyme (Serva, Heidelberg, Germany). The plants were stirred with a plastic tip to ensure an even distribution of the solution and let sit for 4 hours. Protoplasts were released from the seedlings by gently swirling the petri dish for 5 min. The buffer containing the protoplasts was transferred to polypropylene centrifuge tubes. The operation was repeated several times with 10-20 ml of fresh buffer until no further protoplasts were released. The protoplasts were centrifuged for 10 min at 130 g in a swing out rotor, gently resuspended in 5 ml of digestion buffer with a modified pH (pH=6) and centrifuged again for 5 min at 130 g. The protoplasts were resuspended in 5 ml of breakage buffer (300 mM sorbitol, 20 mM Tricine-KOH pH8.4, 5 mM EGTA, 5 mM EDTA, 10 mM NaCO3, 0.1 % PIC and 0.1 % Bovine Serum Albumin (BSA)) and were forced twice through a “Protoplast rupturing device” consisting of two layers of nylon mesh filter (23 and 18 µm mesh) fixed with tape at the top of the cut end of a 10 ml syringe. The broken protoplasts were layered on the top of a two phase density gradient of colloidal silica particles (40 and 85% Percoll (Sigma) in 300 mM sorbitol, 20 mM Tricine-KOH pH 8.5, 2.5 mM EDTA, 5 mM MgCl2 and PIC 0.1% (only in the 40% Percoll phase)) and centrifuged 20 min at 2600 g. The band of intact chloroplasts at the interphase of the two Percoll layers was gently sucked up with a plastic Pasteur pipette, transferred to a polypropylene tune and washed in 40 ml of Hepes-Sorbitol buffer (HS: 330 mM sorbitol, 50 mM HEPES pH 8). The chloroplasts were centrifuged at 700 g for 5 min and re-suspended in 500 µl of HS buffer.

Isolation of P. sativum chloroplasts:

100-150 g of aerial parts of 10-14 day old pea seedlings grown on soil were harvested, rinsed with cold tap water and ground in 400 ml of cold grinding buffer (330 mM Sorbitol, 50 mM Hepes-KOH pH 7.5, 2 mM EDTA, 1 mM $MnCl_2$, 1 mM $MgCl_2$, 5 mM Na-ascorbat and 0.025% BSA) with a rotary homogenizer (Waring commercial blender, Bender and Hobein). The plant extract was filtered through 2 layers of cheesecloth and 2 layers of Miracloth and centrifuged for 2 min at 1600 g. The chloroplasts were gently re-suspended in 5 ml of grinding buffer and loaded on the top of two Percoll gradients (40 and 85% of PBF-Percoll (3% PEG400, 1% BSA, 1% Ficoll in Percoll) solution in 25 mM HEPES-KOH pH7.5, 330 mM sorbitol, 2 mM EDTA, 1 MgCl2 1 mM, 500 µM reduced glutathione and 0.1% sodium ascorbate). The gradients were centrifuged for 15 min at 1400 g and the intact chloroplasts at the interphase of the gradients were carefully sucked up with a Pasteur pipette, washed with 40 ml of cold Hepes-Sorbitol buffer (HS: 330 mM sorbitol and 50 mM HEPES-KOH pH 7.5) and collected by centrifugation for 5 min at 1000 g. Chloroplast were re-suspended in 0.5-1 ml of HS buffer and the chlorophyll concentration was measured.

Chlorophyll measurement:

5-10 µl of chloroplasts were re-suspended in 1 ml of 80% acetone, vortexed and centrifuged at 16'000 g for 3 min. The DO_{652} of the supernatant was measured and divided by the specific extinction coefficient of chlorophyll (36 $L*mg^{-1}*cm^{-1}$) and corrected for the dilution (Arnon, 1949).

In vitro import:

In vitro synthesis of radiolabeled proteins:

Radiolabeled proteins were produced using a rabbit reticulocyte lysate (TNT® Quick Coupled Transcription/Translation Systems, Promega). 40 µl of the reticulocyte lysate were mixed with 1-2 µg of plasmid DNA and 20 µCi of L-[35S] methionine (1000 Ci/mmol) for a total volume of 50 µl. The reaction was incubated at 30°C for 90 min and stopped by placing it on ice.

In vitro import:

Equivalents of 25 µg of chlorophyll of isolated *A. thaliana* chloroplasts in import buffer (0.33 M sorbitol, 0.1 M HEPES-KOH pH7.5, 0.1 M DTT, 0.5 M KOAc and 0.1 M $Mg(OAc)_2$, 0.01 M L-Methionine, 5 mM ATP) for a total volume of 45 µl were incubated in the dark at 25°C for 10 min. The import reaction was started by adding 5 µl of a [^{35}S] Met radiolabeled *in vitro* translated pre-protein in a total volume of 50 µl. Reactions were incubated at 25°C in the light and were stopped by adding 1 ml of cold HS. Chloroplast were collected by centrifugation for 1 min at 800 g. Chloroplasts were carefully re-suspended in 1 ml of cold HS and centrifuged again. The chloroplast pellet was frozen or used for subsequent protease treatment or fractionation.

Protease treatment (Froehlich, 2011):

The re-isolated chloroplast pellet of an import reaction was re-suspended in 100 µl of HS buffer and thermolysin or trypsin was added to a final concentration of 5 or 50 µg/ml. The reaction was incubated on ice for 30 min. Thermolysin treatment was stopped by diluting the reaction in 1 ml of 10 mM EDTA in HS. The trypsin reaction was stopped by adding trypsin inhibitor to a final concentration of 330 µg/ml. Chloroplast collected by centrifugation for 1 min at 0.8 g. The pellet was re-suspended directly in SDS-PAGE sample buffer.

Chloroplast fractionation and alkaline treatment:

The re-isolated chloroplast pellet of an import reaction was resuspended in 150 µl of cold HS and then diluted in 800 µl of cold 2 mM EDTA. The tubes were vortexed and incubated on ice for 10 min. The membrane and soluble fractions were separated by centrifugation for 30 min at 16'000 g. The supernatant was collected for protein precipitation. The pellet was re-suspended in 1 ml of cold $NaCO_3$ 0.1 M pH 11.5 using a plastic Pasteur pipet. The reaction was incubated on ice for 10 min. The membranes were collected by centrifugation at 40 000 g for 30 min. Proteins of the soluble fraction were precipitated by the MeOH/$CHCl_3$ extraction methods while the membranes pellet was directly dissolved in the appropriate volume of sample buffer.

TAP-Toc159 purifications:

When possible, manipulations were carried out in a cold room at 4°C. 10 g of three weeks old seedlings grown on agarose medium were ground in a cold mortar in 15 ml of extraction buffer containing protease inhibitors (50 mM Tris-HCl pH 7.5, 100 mM NaCl; Inhibitors: 1 mM PMSF, 0.2 % Protease Inhibitor Cocktail for plants (PIC, Sigma-Aldrich)). The extract was filtered through two layers of Miracloth (Millipore) and the volume was adjusted to 35 ml. The filtered extract was centrifuged twice for 5 min at 1'500g. The supernatant was centrifuged or 1h at 100 000 g in a SW32-Ti rotor in an OPTIMA™ XPN ultracentrifuge (Beckman-Coulter). The membrane fraction was resuspended in extraction buffer using a Potter homogenizer. The resuspended membrane fraction and the soluble fraction were centrifuged for 1h at 100 000 g. The pellet of the membrane fraction was solubilized in extraction buffer containing 0.75% TX-100 and 10% glycerol. The solubilized membrane fraction and the soluble fraction were incubated over night with IgG sepharose equilibrated with extraction buffer (10 µl of beads/g starting material; HsIgG, MP Biomedicals, Irvine, USA; CNBr-activated sepharose 4B, GE Healthcare). The beads were washed once with the same volume of extraction buffer as the load, and 5 times with 50 bed volumes. The last wash was done with extraction buffer without proteases inhibitors. The beads were transferred to a 0.5 ml spin column and eluted for 2 hours at 16°C with AcTEV protease (Tobacco Etch Virus protease, Invitrogen) according to the provider's instructions, except that 100 mM NaCl and TX-100 0.75% were added to the beads that had been incubated with the membrane fraction. The spin columns were centrifuged for 1 min at 100 g to collect the eluate supernatants which were flashed frozen in liquid nitrogen and stored at -80°C.

Emb2004-TAP purification:

Chloroplasts equivalent to 2 mg of chlorophyll were solubilized for 30 min by gentle rotation in extraction buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl; 0.75% TX-100, 10% Glycerol; Inhibitors: 1 mM PMSF, 0.2 % Protease Inhibitors Cocktail for plant (PIC, Sigma-Aldrich)) at a final concentration of 0.5 mg/ml of chlorophyll. The solubilized extract was centrifuged for 1h at 100 000 g, in a SW55-Ti rotor (Beckman-Coulter). The supernatant was run repeatedly for a total duration of 2h over a chromatography column packed with 100 µl of IgG sepharose. The resin was washed once with the same volume of extraction buffer as the load, and 5 times with 50 bed volumes. The resin was transferred to a 0.5 ml spin-column and eluted for 10 min under gentle rotation using 200 µl of 0.1 M Glycine pH 3.0. The spin-column was centrifuged 1 min at 100 g to collect the glycine eluate supernatant. The eluate was flash frozen in liquid nitrogen and stored at -80°C.

Stable and transient transformation:

A. thaliana transformation by floral Dip:

Transformation by *A. tumefaciens* and Floral Dip of *A. thaliana* were adapted from "Arabidopsis A Laboratory Manual" (D. Weigel, J. Glazebrook, CHSL PRESS, 2001). A *A. tumefaciens* 5 ml pre-culture containing the T-DNA to be inserted was used to inoculate a 200 ml culture that was incubated for approx. 24h at 28°C until >2 OD_{600} was reached. The bacteria were collected by centrifugation at 3000 g and re-suspended in dipping solution to get a final OD_{600} of 0.2 (0.22% MS salt with vitamins, 5% sucrose, 1µM acetosyringone, 0.05% v/v Silwet L-77). The inflorescences of *A. thaliana* plants were dipped for 30 sec in the *A. tumefaciens* suspension. The plants were laid down in a plastic tray and enclosed in a black bin bag to retain humidity and protect bacteria from the light for 3 days. Mature seeds were sown on soil and selected for resistance by spraying the seedlings with a 300 µM glufosinate ammonium solution every 3-4 days. Resistant plants were transferred to individual pots for further characterization.

Agro-infiltration of N. benthamiana:

Agro-infiltration of *N. benthamiana* was done according to Wroblewski et al. (2005). A culture of *A. tumefaciens* carrying the T-DNA plasmid containing the sequence of interest was grown in LB liquid medium at 28°C for 2-3 days. Cells were collected by centrifugation and re-suspended in agro-infiltration solution (10 mM MES pH 5.7, 150 µM acetosyringone and 10 mM $MgCl_2$) to reach an OD_{600} of 0.4. Using a 10 ml syringe, the cell suspension was gently infiltrated in *N. benthamiana* leaves. The expression of the recombinant fluorescent protein was observed on whole mount leaf preparations by laser scanning confocal microscopy using a LEICA SP2 AOBS microscope SP5 confocal microscope (Leica, Wetzlar, Germany). The excitation wavelength for the YFP and the chlorophyll signals were 514 nm and 594 nm, respectively. The windows of detection for the emission signals of YFP and chlorophyll were between 520-588 nm and 680-514 nm, respectively.

Protoplast preparation and PEG-transformation:

A method adapted from Jin et al. (2001) was used. Approx. 5 g of 4 weeks old *A. thaliana* plants were incubated in a petri dish over night with enzyme solution (400 mM mannitol, 5 mM MES-KOH pH 5.6, 8 mM CaCl2, 1% Cellulase Onozuka R-10 and Macerozyme R-10 (Serva). Protoplasts were released by gently shaking the petri dish. Protoplasts were filtered through a 100 µm mesh nylon filter, collected by centrifugation for 5 min at 100 g and re-suspended in 10-15 ml of W5 solution (154 mM NaCl, 125 mM CaCl2, 5 mM KCl, 5 mM glucose, 5 mM MES-KOH pH 5.6). The concentration of protoplasts was evaluated with a counting chamber. Protoplasts were collected by centrifugation for 5 min at 100 g and re-suspended to reach a final concentration of 1.5 10^6 protoplasts/ml in MaMG solution (0.4 M mannitol, 15 mM MgCl2, 5 mM MES-KOH pH 5.6). 20 µg of plasmid DNA was gently mixed in a 2 ml centrifuge tube with 100 µl of the protoplast solution and 110 µl of PEG solution (40% PEG4000, 0.4 M mannitol, 0.1 M $Ca(NO_3)_2$, 0.1% MES-KOH pH 7-8) to reach a final volume of 220 µl. The reaction was incubated for 20 min at room temperature and washed in 1.5 ml of W5 by gently inverting the tube. Protoplasts were collected by centrifugation for 5 min at 100 g and washed in 1 ml of protoplast-culture medium (4.4 mg/ml MS salts, 350 mM sorbitol, 50 mM Glucose, 3 mM $CaCl_2$, 0.1 mg/ml ampicillin, pH 5.8 (KOH)). Protoplasts were pelleted down at 100 g for 2 min, resuspended in 500 µl of protoplast-culture medium and incubated for 24-48h in the dark before observation by confocal microscopy. All the manipulations were done under a sterile hood using sterile solutions and plastic ware.

Proteins methods:

Protein extractions:

50 mg of leaves were ground in a 1.5 ml microfuge tube with a plastic pestle in 400 µl of cold grinding buffer (100 mM NaCl, Tris-HCl pH 7.5, 0.5% TX-100, 10 mM β-mercaptoethanol, 1 mM PMSF, 0.2% PIC, Rensink, 1998). The ground material was incubated under gentle rotation for 30 min at 4°C and then centrifuged for 10 min at 16'000 g. 200 µl of the supernatant were precipitated by $CHCl_3$/MeOH precipitation.

Protein quantification:
Protein concentrations in plant extracts were measured using the Bradford reagent (Bio-Rad Protein Assay Dye Reagent Concentrate, BIO-RAD). The OD_{595} was determined and the concentration calculated and corrected for dilution with a specific extinction coefficient of 0.055 $L*cm^{-1}*ug^{-1}*ml$. The coefficient was calculated using BSA as a standard.

Protein precipitation by chloroform/methanol (CHCl3/MeOH precipitation):
2.4 v. of methanol and 0.8 v. of chloroform were added to 1 v. of protein extract and mixed by vortexing. 3.2 v. of H_2O were added resulting in a white precipitate. After vortexing, the extract was centrifuged for 1 min at 16'000 g. The upper phase just above the white precipitate was discarded, the lower phase and the precipitate were vortexed with 2.4 v. of methanol and centrifuged at 16'000 g for 5 min. The supernatant was removed and the pellet was dried in the open air or in a vacuum concentrator. The dried pellet was re-suspended in the appropriate volume of sample buffer (SB: 0.05 M Tris-HCl pH 6.8, 0.1 M Dithiotreitol (DTT), 2% sodium dodecyl sulphate (SDS), 0.1% Bromophenol blue, 10% glycerol) and incubated for 10 min at 65°C to dissolve the proteins.

SDS-PAGE:
Proteins dissolved in sample buffer were separated by SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) according to "Molecular Cloning a Laboratory Manual" (Sambrook and Russel, CSHL PRESS, 2001). Gels were run using Mini-PROTEAN® (BIO-RAD) or "PerfectBlue Dual Gel System Twin" (peqlab) electrophoresis systems. The protein standard used for SDS-PAGE was the "unstained broad range standard" (6.5-200 kDa, Ref. #161-0317) from Bio-Rad. After separation, gels were either dried using a "Gel Dryers" (Model 583, Bio-Rad) or transferred on nitrocellulose membrane by western blotting.

Western Blotting:

Western blotting was done according to "Molecular Cloning a Laboratory Manual" (Sambrook and Russel, CSHL PRESSS, 2001), using a Trans-Blot® transfer cell (BIO-RAD) and the DUNN carbonate transfer buffer.

Yeast:

Yeast media:

YPDA: 20 g/l Difco peptone, 10 g/l Yeast extract, (2% bacteriological agar for solid media). 0.003% adenine hemisulfate and 2% glucose were added after autoclaving (121°C, 1 atm, 20 min).

Drop-Out (DO) medium: 2.7 g/l of yeast nitrogen base without amino acids (Becton, Dickinson *and Company*) (20 g/l agar for bacteria added for solid medium) and the appropriate 10X drop-out solution (L-Adenine hemisulfate salt :200 mg/l; L-Arginine HCl 200 mg/l; L-Histidine HCl monohydrate

200 mg/l; L-Isoleucine 300 mg/l; L-Leucine 1000 mg/l; L-Lysine HCl 300 mg/l; L-Methionine 200 mg/l; L-Phenylalanine 500 mg/l; L-Threonine 2000 mg/l; L-Tryptophan 200 mg/l; L-Tyrosine 300 mg/l; L-Uracil 200 mg/l; L-Valine 1500 mg/l).

Yeast was grown at 30°C with a 250 rpm agitation when in liquid medium.

Yeast transformation:
Yeast strains were transformed according to the "LiAc mediated yeast transformation" protocol of the "Yeast Protocols Handbook" (Clontech, PT3024-1, March 2001) based on the LiAc method (Ito, et al, 1983) modified by Schiestl and Gietz (1989), Hill *et al*. (1991) and Gietz *et al* (1992).

Yeast protein extracts:

10 ml cultures of the appropriate DO medium inoculated with a fresh colony were grown over-night at 30°C at 250 rpm. The OD_{600} was measured and a new culture was inoculated in YPD to have a starting OD_{600} of 0.2. Yeasts were incubated 3-4 at 30°C, 250 rpm to reach a final OD_{600} of 0.6. 3 units of OD600 were centrifuged at 16 000 g for 1 min. The pellet was re-suspended in 300 µl of 50 mM phosphate buffer (K_2HPO_4-KH_2PO_4) pH 7.4 and 100 µl of TCA were added and the extract was incubated for at least 30 min at -80 °C to let the protein precipitate. The extracts were centrifuged at 4°C and 16'000 g for 10 minutes. The pellet was washed twice with 80% acetone and then dried using a vacuum concentrator. The pellet was re-suspended in the appropriate volume of 0.1 M NaOH and 1% SDS. 3 volumes of 4X sample buffer were added and the tubes were incubated at 99°C for 10 min.

Pairwise yeast two-hybrid assays:
The pGADT7 and pGBKT7 vectors containing the sequences to be assayed were co-transformed in the AH109 yeast strain. Positive clones were plated on Drop-Out medium lacking leucine and tryptophan for the selection of the pGADT7 and pGBKT7 vector respectively.

Yeast two-hybrid screen:
The Matchmaker Construction and Screening Kit (Clontech, TAKARA) was used. It provided all the plasmid, primers, enzymes and reagents needed for the screen (the solid and liquid media were prepared in-house).

Primers for the first strand synthesis of the cDNA library:

SMART III Oligo: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCGGG-3'
CDS III Primer: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d$(T)_{30}$VN-3'
CDS/6 Primer: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-NNNNNN-3'
(V=A/G/C; N=A/G/C/T);

Primers for the amplification of the library:

Forward: 5'TTCCACCCAAGCAGTGGTATCAACGCAGTGG-3'
Reverse: 5'-GTATCGATGCCCACCCTCTAGAGGCCGAGGCGGCCGACA-3'

Reverse transcriptase: MMLV reverse transcriptase (Moloney Murine Leukemia Virus);

Plasmids: pGADT7-Rec for the cloning of the library; pGBKT7-53 Control Vector; pGBKT7-Lam control vector; SV40 Large-T PCR fragment.

Hybrigenics Yeast Two-Hybrid screen:
The sequence of Emb2004 lacking the transit peptide and the transmembrane domain (Lys11-Phe605) was ligated at the C-terminal of the DNA-binding domain of *E. coli* LexA and was screened against a cDNA library (cloned in the pP6 vector to give a C-terminal fusion to the Gal4 activating domain) of one week old *A. thaliana* seedlings. Positive interactions were selected for the activation of the *HIS3* reporter gene.

DNA cloning, genotyping TAIL PCR and qPCR:

Cloning:

Plasmids:

pET21d-Emb2004: The Emb2004 full length sequence was amplified from the Emb2004 cDNA plasmid obtained from the ABRC stock (ref: u11104) with primers adding an EcoRI site (5'-CTATAGAATTCATGGCTTCTTCTTCCACC-3') and a STOP codon and a NotI site (5'-GATATGCGGCCGCCTAGAAGAAAATTTCTACTTCC-3'). The insert was cloned in pET21d in EcoRI/NotI.

pET21d-AT4G32250: The AT4G32250 full length sequence was amplified from the AT4G32250 cDNA obtained at the ABRC stock (ref: U18979) with primers adding a Nco1 site (5'-CTATACCATGGTGGCTGTTCAG-3') and a STOP codon and a NotI site (5'-GATATGCGGCCGCTCATAGTGTGATAACTTGG-3'). The insert was ligated into pET21d using Nco1/NotI.

pETM40-Emb2004: The Emb2004 sequence without the transit peptide and the transmembrane domain (61ala-605phe) was amplified from the pET21d-Emb2004 plasmid with 5'-CTATAGGCCGGCCGAATTCATGTCGCCTAAGTCTCC-3' adding an EcoRI site and 5'-GATATGCGGCCGCCTAGAAGAAAATTTCTACTTCC-3' adding a STOP codon and a NotI site. The PCR product was ligated into the EcoRI/NotI sites of pETM40.

pEG205-Emb2004 and pEG101-Emb2004: The Emb2004 full length sequence was amplified from the pET21d-Emb2004 plasmid with primers adding attB recombination sequences in 5' and 3' of the insert. (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCTTCTTCTTCCACCAG-3'
5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGGAAGAAAATTTCTACTTCC-3'). The insert was cloned in the GATEWAY pDONR221® vector by recombination, according to the provider's instructions (life technologies, Paisley, UK). The Emb2004 full length sequence was ligated into the pEARLYGATE205 and pEARLYGATE101 vector by recombination to obtain a C-terminal fusion with the TAP tag (Tandem Affinity Purification tag) or YFP (Yellow Fluorescent Protein), respectively.

pEGAT4G32250: AT4G32250 full length sequence was cloned the same way as Emb2004 in pDONR and pEARLYGATE101 with the following primers: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGTGGCTGTTCAGTTGT-3'
5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGTAGTGTGATAACTTGGTCT-3'.

Plasmids for Yeast Two-Hybrid assays:

The following sequences were ligated into the pGBKT7 and the pGADT7 vectors at C-terminal of the Gal4 binding or activating domains, respectively: Toc159GM (F854 -Y1503t), Toc159G (A846-V1093), Toc159M (F1102-Y1503), Toc33 without the N-terminal transmembrane domain (32 AA) (M1-K265), Toc75 without the transit peptide (141asp-818tyr), Emb2004 without the transit peptide (A61-F605), Emb2004 without the transit peptide and the transmembrane domain (F116phe-F605.

Bacteria media and growth conditions:

LB according to Miller *autoclaved at* (solid media: 0.8% agar bacteriological grade) was used to grow *E.coli* (DH5α, BL21) and *A. tumefaciens* (C58). The antibiotics were added after autoclaving. Concentration were as follows: ampicillin: 100 µg/ml; kanamycin: 50 µg/ml; rifampicin: 50 µg/ml. *E. coli* (DH5α, BL21) and *A. tumefaciens* bacteria were grown at 37°C and 28°C respectively, with agitation at 150 rpm.

In vitro DNA recombination

Restriction enzymes were all purchased from *New England Biolabs®* Inc. (NEB). The competent DH5α or BL21 cells were prepared according to Inoue et al. (1990).
The PCR fragment to be cloned was amplified from c.a. 100 ng of a "donor" plasmid with 5 units for a 50 µl reaction of the proof-reading PWO polymerase (Roche Diagnostics, Germany) with an extension time of 1 min/kB at 72°C and an annealing temperature adapted to the pair of primers used. The PCR product and the appropriate plasmid were digested with the chosen restriction enzymes (NEB), separated on a 1% agarose (Standard agarose-Type LE) gel. The appropriate bands were excised from the gel and extracted using a DNA gel extraction kit (Wizard®Plus SV Gel and PCR Clean-Up System, Promega). The 5'-phosphate of the digested PCR insert was removed by phosphatase treatment (Antarctic Phosphatase, NEB). The ligation of the plasmid and the PCR insert was done using T4 DNA ligase (NEB).

Transformation in E. coli

Ligation product or plasmids were transformed in competent *DH5α E. coli* according to the method in "Molecular Cloning, a Laboratory Manual 2 (Sambrook and Russel, CHSL PRESS, 2001) except that the SOC medium was replaced by LB medium. The transformed bacteria were plated on the appropriate selective LB plate (ampicillin: 100 µg/ml; kanamycin: 50 µg/ml). The positive clones were analyzed by PCR and the plasmids were prepared according to a modified version of the alkaline lysis method (Birnboim & Doly 1979) and verification by restriction digest. Positive plasmids were purified using plasmid purification kit (GenElute™ Plasmid Mini/Maxi prep Kit, Sigma-Aldrich) and sent for sequencing.

Electroporation of plasmid DNA in A. tumefaciens:

T-DNA vectors were transformed in *A. tumefaciens* C58 strain by electroporation, using a "Micropulser" electroporator (Bio-Rad). The transformed bacteria were plated on LB agar medium selective for the T-DNA plasmid *A. tumefaciens* (rifampicin 50 µg/ml) and the Ti plasmid (ampicillin 100 µg/ml) and incubated at 28°C. The proper transformation of the T-DNA was controlled by PCR on the positive colonies.

Rapid DNA preparation (Edwards, 1991):

Approx. 100 mg of plant leaf material was ground in a 1.5 ml microfuge tube with a plastic pestle in 400 µl of extraction buffer (0.2 M Tris-HCl pH 9, 0.4 M LiCl, 25 mM EDTA, 1% SDS), vortexed for 5 seconds and left for 1 h at room temperature under gentle agitation. The extract was centrifuged at 16'000 g for 5 min, 300 µl of the supernatant was transferred in a new microfuge tube and precipitated with 2.5 v. of 100% ethanol. The DNA was pelleted by centrifugation for 5 min at 16 000 g. The DNA pellet was washed with 70% ethanol and briefly air dried. The pellet was re-suspended in 50 µl of H_20.

CTAB (cetyltrimethylammonium bromid) DNA extraction:

100 mg of plant leaf tissue were harvested and frozen in liquid nitrogen in a 2 ml microfuge tube containing a steel bead. The tissue was pulverized using a bead homogenizer. Pre-heated 2X CTAB extraction buffer (2% CTAB, 100 mM Tris-HCL pH 8, 20 mM EDTA, 1.4 M NaCl, 1% polyvinylpyrrolidone (PVPP)) corresponding to 2 vol. of the estimated volume of ground tissue was added followed by incubation for 20 min at 65%. A volume of 24:1 chloroform : isoamyl alcohol equal to the total volume of the extract was added. The tubes were vortexed until only one phase was observed. The tubes were centrifuged at 16 000 g for 1 min. The supernatant was transferred to a new tube, 1/5 of the pipetted volume of 5% CTAB was added (5% CTAB, 0.7 NaCl) and the extract was vortexed. The chloroform : isoamyl alcohol extraction was repeated once, and 1 vol. of precipitation buffer was added (1% CTAB, 50 mM Tris-HCl pH 8 and 10 mM EDTA). The tubes were centrifuged 5 min at 16 000 g and the supernatant was removed. The pellet was re-suspended in 100 µl of high salt TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA and 1 M NaCl) and 250 µl of 100% ethanol was added. The tube was centrifuged 5 min at 16 000 g and the DNA pellet was washed with 1 ml of 70% ethanol. The pellet was air dried for 30 min and rehydrated in 50 µl of H_2O.

Genotyping:

DNA extracted with the "Rapid DNA preparation" method was used.

The wild type allele was amplified with primers flanking the T-DNA insertion of the *emb2004-1* mutant (5'-TCATGTAATTCGGTGATTGC-3' ; 5'-GACAAAAGGAAGATGGCACT-3') resulting in a 405 bp PCR product.

The mutant *emb2004-1* allele was amplified with a primer in the left border of the T-DNA insertion (5'- TAGCATCTGAATTTCATAACCAATCTCGATACAC-3') and in the genomic sequence (5'-GCACCCAAAGCACCACC-3') resulting in an approx. 1500 bp PCR product.

The Emb2004-TAP allele was amplified with a prime in the 35S promoter (5'-TCTCCACTGACGTAAGGGAT-3') and in the Emb2004-TAP coding sequence (5'-GCACCCAAAGCACCACC-3') resulting in an 1108 bp product.

PCR were done using GoTaq polymerase according to the provider's instruction (Promega). The PCR products were separated on 1.5% agarose gel and stained with Ethidium Bromide.

Thermal Interlaced PCR (TAIL PCR):

The TAIL PCRs for the genomic localisation of T-DNA insertions were done according to "ARABIDOPSIS, a Laboratory manual (D. Weigel, J. Glazebrook, CHSL PRESS, 2002) except that right and left border primers were designed specifically for the pEG205 plasmid.

RB1: 5'-CGAACTAATAACGCTCACTGAAGGGAACTC-3'; RB2: 5'-GCATGGGTGAGATTCCTTGAAGTTGAGTAT-3'; RB3: 5'-CAAATGAAGTGCAGGTCAAACCTTGACAGT-3';
LB1: 5'- TATTCAGTAATCTCGGCCAATATCCTAAATGTG-3' LB2: 5'-TAAATGTGCGTGGCTTTATCTGTCTTTGTATTGT-3'; LB3 5'-TTCATCAATTCATGTAACGTTTGCTTTTCTTATG-3'.

Degenerate primers:
5'-NGTCGASWGANAWGAA-3' 128-fold degenerate
5'- TGWGNAGSANCASAGA-3' 128-fold degenerate
5'-AGWGNAGWANCAWAGG-3' 128-fold degenerate

5’- STTGNTASTNCTNTGC-3’ 256-fold degenerate
5’-NTCGASTWTSGWGTT-3’ 64-fold degenerate
5’-WGTGNAGWANCANAGA
S = G/C; W = A/T; N = A/C/G/T

The DNA used for the TAIL PCR was extracted from leaves of EMB2004-TAP n°11 line plants with the CTAP DNA extraction method. GoTaq® DNA polymerase from Promega was used.

RNA extraction and qRT-PCR:

RNA from 20-100 mg of a plant leaf was ground with mortar and pestle in liquid nitrogen. The RNA was extracted with a commercial kit combining guanidine-isothiocyanate lysis and silica-membrane purification (NucleoSpin RNA, Macherey-Nagel AG, Düren, Germany). The RNA was quantified by spectrophotometry (260 nm) and was reverse transcribed with the GoScript reverse transcriptase (Promega, Madison, WI, USA) using a 15mer Oligo(dT) primer for priming. The cDNAs were flash frozen in liquid nitrogen and stored at -80°C.

qPCR was done with a relative quantification method, using the non-specific dye SYBRgreen (FastStart Essential DNA Green Master, Roche, Basel).

Primers:
native *Em2004* cDNAs: 5’-TGAAGGCCAATGAAGATG-3’; 5’- ATGGTGCAAAATTACAGAGTGA-3’
Emb2004-TAP cDNAs: 5’-TGCGCTCACGGATGCTA-3’; 5’-GCGGTTGGCTGCTGAGAC-3’

Control genes:

AT1G13320: 5’-TAACGTGGCCAAAATGATGC-3’; 5’-GTTCTCCACAACCGCTTGGT-3’
AT4G26410: 5’-GAGCTGAAGTGGCTTCCATGAC-3’; 5’-GGTCCGACATACCCATGATCC-3’
AT4G27960: 5’-TCACAATTTCCAAGGTGCTGC -3’; 5’-TCATCTGGGTTTGGATCCGT-3’

All the primer pairs were checked for efficiency between 95 and 105% on 10^{-1} serial dilutions of cDNA. The qPCR were done on a “qPCR LightCycler 96” (Roche, Basel) with 45 (15 s - 95°C; 30 s - 60°C; 30 s – 72°C) cycles. The total reaction volume was of 20 µl, consisting of 10 µl of qPCR mix, 100 nM each primers and 0.1 µl cDNA. The ratio of expression between the samples was calculated by the ΔΔCt method (Livak): The C_T of the gene of interest is normalized by the C_T of the reference gene in both the test sample and the calibrator sample (C_T being the number of cycles at which the SYBRgreen fluorescence of a reaction passes a threshold).

$\Delta C_{T(test)} = C_{T(target,test)} - C_{T(ref,\ test)}$
$\Delta C_{T(calibrator)} = C_{T(target,\ calibrator)} - C_{T(ref,\ calibrator)}$

Then, the ΔC_T of the test sample is normalized by the ΔC_T of the calibrator

$\Delta\Delta C_T = C_{T(test)} - C_{T(calibrator)}.$

The ratio of expression of the gene of interest between the test sample and the calibrator sample is given by: $2^{-\Delta\Delta Ct}$

Microscopy:

Electron microscopy:

Small leaf pieces of about 10 mm^2 were vacuum infiltrated with a fixation solution (4% formaldehyde, 5% glutaraldehyde, 0.1 M $NaHPO_4/Na_2PO_4$ buffer pH 6.8 (NaPi)) and incubated over night at 4°C under gentle agitation. The fixed material was washed three times with NaPi and post-fixed in 1% OsO_4 in NaPi for 2h. The samples were dehydrated by successive incubations in increasing concentrations of ethanol (30%, 50%, 70%, and 90%) and acetone (90% and 100%). The samples were exposed to increasing concentration of low viscosity Spurr's resin in acetone (33%, 66%, 100%; Spurr's resin: 0.83% (w/w) dimethylaminoethanol, 34% (w/w) vinylcyclohexenedioxine, 15.83% (w/w) Diglycidylether polypropylene glycerol, 49.16% (w/w) Nonenyl succinic anhydride)(Spurr, 1969). The resin was polymerized over-night at 60°C. 50-70 nm ultrathin sections were cut from the resin block with a diamond knife and fixed on uncoated copper grids. The sections were contrasted with uranyl acetate and Reynold's lead citrate solution and observed with a transmission electron microscope (Philips CM 100, Philips Electron Optics BV, Eindhoven, the Netherlands).

Light microscopy:

750 nm sections from resin embedded material (described in Electron microscopy section) were prepared with a diamond knife. Sections were stained with toluidine blue (0.6% Toluidine Blue, 70 mM Na_2CO_3, 7% ethanol) for 2 min at 50°C, assembled on a glass-slide and observed with a Nikon Eclipse E800 light microscope (Nikon AG, Egg, CH).

Mass-spectrometry:

SyproRuby staining:

A SDS-PAGE gel was fixed in a 50% methanol, 7% acetic acid solution for 30 min twice. The gel was stained over-night with SYPRORuby and washed for 30 min with a 10% methanol and 7% acetic acid.

TAP-Toc159 TEV eluate analysis:

The proteins of the TAP-Toc159 TEV eluate were separated on gradient NuPAGE Novex BisTris gels (4% to 12%), and stained with SyproRuby (Invitrogen). The gel was cut in 3 bands for separate analysis. The trypsin in-gel digestion with was done according to a modified protocol from Shevchenko et al. (1996, 2006). The samples were desalted before being analyzed by mass-spectrometry. The dried peptides were resuspended in 3% (v/v) acetonitrile (ACN), 0.2% (v/v) formic acid and analyzed on a LTQ FT-ICR mass spectrometer (Thermo Fisher Scientific) coupled with an Eksigent Nano-liquid chromatography system (Eksigent Technologies). The peptide mixtures were loaded onto a chromatography column packed with Magic C18 AQ beads. The peptides were eluted from the column by an increased ACN concentration in the mobile phase from 5% to 40%. The tandem mass spectrometry spectra were searched against the "Arabidopsis Information Resource 8 protein database" (downloaded on December 14, 2007) using the Mascot 2.1.04 software (Matrix Science).

Emb2004-TAP glycine eluate analysis:

The EMB2004-TAP glycine eluate was run on a SDS-PAGE gel and lanes were divided in 5 bands for separate analysis. The preparation and in gel digestion of the gel was done according to a protocol

for the *P. pastoris* recombinant proteomic grade Trypsin (Cat. No. 03 708 985 001, Roche, Basel). Briefly, the lanes of the SyproRuby stained gel were cut into bands for separate analysis. The bands were diced into small pieces (around 1 mm^3), distained with a 30% acetonitrile 100 mM NH_4CO_3 solution washed in 100% acetonitrile and dried. The proteins in the samples were treated with a reducing solution (10 mM DTT and 100 mM NH_4HCO_3), alkylated with a 54 mM iodoacetamid, 100 mM NH_4HCO_3 solution, destained with 30% acetonitrile in 100 mM NH_4HCO_3 and dried. The proteins were digested over-night in a 0.01 mg/ml Trypsin solution. Peptides were extracted in a 35% acetonitrile 0.4% trifluoroacetic (TFA) solution. The samples were desalted and run on a LTQ XL™ Linear Ion Trap Mass Spectrometer.

Bioinformatic:

Prediction and Annotation:

The transmembrane prediction was done with the TMPred algorithm and the transit peptide prediction with ChloroP. The MapManBin category was retrieved from the Plant Proteome Data Base (PPDB). The curated locations of the proteins identified by mass spectrometry were retrieved from the At_Chloro database.

Structure homology modelling:

The tertiary structure of Emb2004 was established by "structure homology modelling" using the SWISSMODEL server which provides a fully automated service. First, a suitable template with a known crystallographic structure is identified based on sequence similarity. The tertiary structure of the query protein is modeled by fitting its sequence to the known crystallographic structure. The template of the porcine Ribonuclease inhibitor protein was identified (template ref: 1z7x.1.B), with a 72% coverage (K153 to the E595) and a sequence similarity of 29%.

Results:

TAP-Toc159 purification for MS analysis:

Purification and western blotting:

To identify new potential proteins interacting with Toc159, a recombinant TAP-Toc159 was purified and the eluate was analyzed by mass spectrometry. A plant line expressing in a *ppi2* background a recombinant Toc159 protein fused in N-terminal with a TAP tag was used (fig. 8, a). The TAP tag consists of two successive proteins A (ProtA) domains fused in C-terminal to a calmodulin binding protein (CBP). The ProtA domains specifically interact with the heavy chain of IgGs, allowing the precipitation of the recombinant protein with IgG beads. The ProtA domains and the CBP are separated by a Tobacco Etch Virus protease (TEV) cleavage site to allow for the elution of the recombinant protein once attached to IgG beads. The TAP-Toc159 construct complements the *ppi2* mutation (fig.8, b). A TAP construct expressed in a wt background was used as a control line. The plant material was ground in a cold mortar in a non-denaturing buffer. The membranes and soluble fraction were separated by ultracentrifugation. After re-suspension and solubilization with the detergent Triton X-100 of the membranes, both fractions were incubated with IgG coated sepharose beads. IgG bind specifically protein A domain. The beads were eluted by TEV protease. Samples of the load, the flow-through and the eluates of the purification were analyzed by SDS-PAGE and Western blotting. The membrane was probed with IgG to detect the recombinant proteins, antibodies against Toc/Tic components and against the Light Harvesting Protein (LHCP), the Large Subunit of the Rubisco (LSU) and actin (fig. 9).

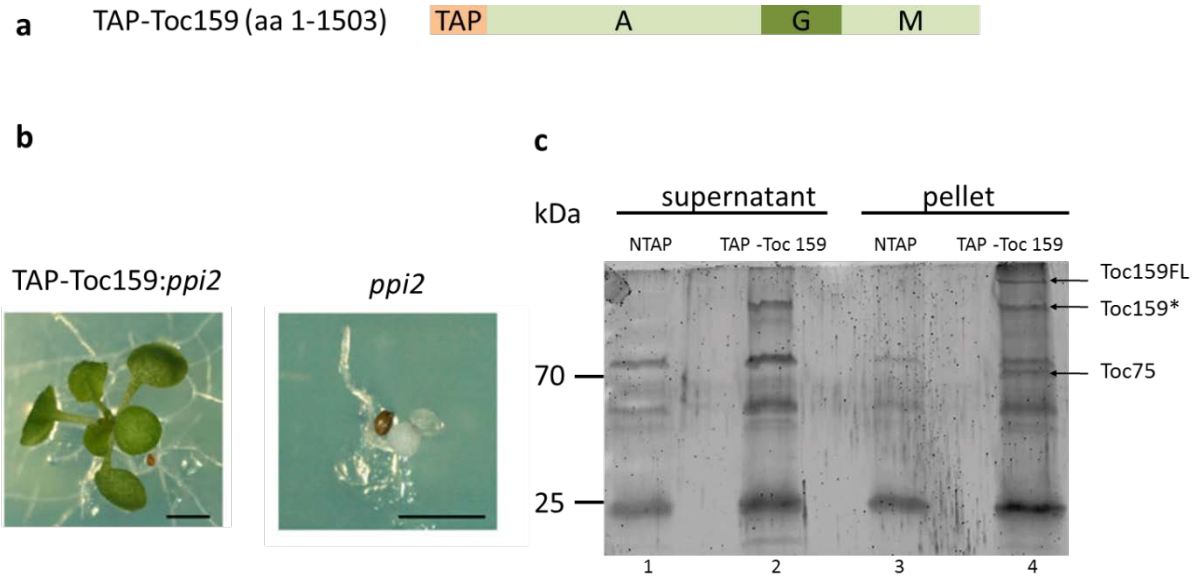


Figure 8 : Purification of a TAP-Toc159 fusion protein. A) Scheme of a construct coding for an N-terminal TAP-Toc159 fusion protein. B) Transgenic plant expressing the stably integrated TAP-Toc159 construct in a *toc159* mutant background (TAP-Toc159:*ppi*2) and the *ppi2* mutant. Bars = 2 mm. C) The TAP-Toc159 recombinant protein was purified from 3 week old seedlings. The membranes were solubilized with 0.75% TX-100 after separation from the soluble fraction. IgG beads were used to isolate the fusion protein from both of the fractions and were eluted with TEV protease. Plants expressing the TAP-tag alone in a wild type background were used as a control. The eluates were loaded on a SDS-PAGE gel and stained with SYPRO ruby. Each lane was cut into three parts which were processed independently. The proteins were tryptically digested in gel, extracted and analyzed by tandem mass-spectrometry. * Proteolytic fragment of TAP-Toc159.

The NTAP was correctly expressed as it was detected in the load of the soluble fraction. It was not detected in the load of the membrane fraction as TAP is a soluble protein. The TAP-Toc159 was identified in both the soluble and the membrane fraction. In the membrane fraction two bands were found, one above 200 kDa corresponding to the full length protein and a smaller fragment between 200 and 116 kDa. The abnormal migration of the full length protein above 200 kDa is expected, as similar migration is known for the wild type Toc159. This is likely due to the highly negative charge of this protein, due to its abundance of acidic residues in the N-terminal acidic A-domain. This can impair the interaction of the protein with the negatively charged SDS and alter the migration pattern. In the eluate of the soluble fraction, only the small Toc159 fragment was identified. Two other Toc proteins, Toc75 and Toc33 were found in the eluate of the membrane fraction, showing that the Toc core complex was efficiently purified. As expected, Toc33 and Toc75 where not found in the eluate of the soluble fraction. Tic110 was identified neither in the soluble nor the membrane fraction. Actin, the Large Subunit of Rubisco (LSU) and the Light Harvesting Protein used as controls for unspecific binding were not found in any of the Toc159 eluates.

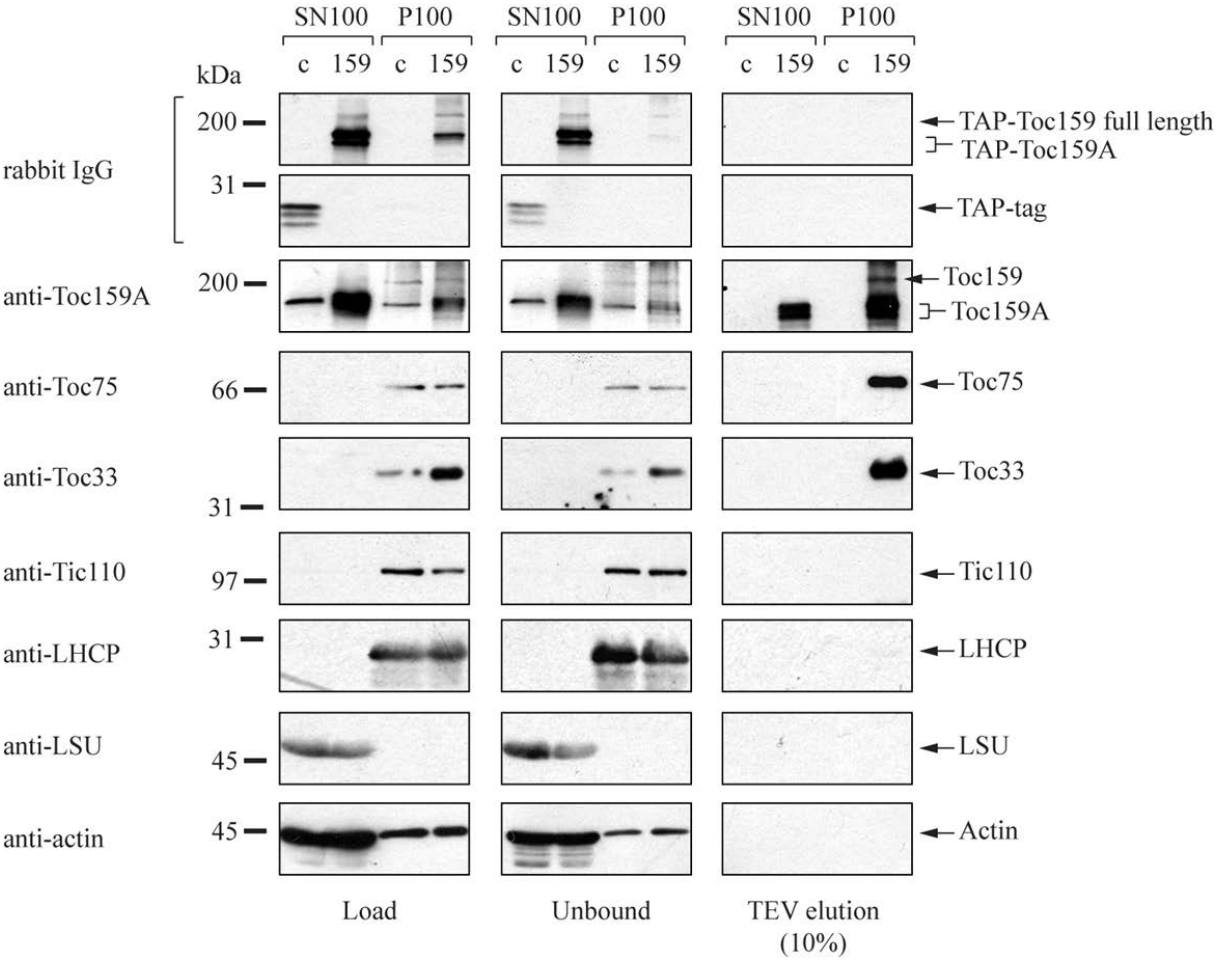


Figure 9 : Toc75 and Toc33 co-precipitate with TAP-Toc159: *A.thaliana* plants expressing a TAP-Toc159 recombinant protein in the *ppi2* background were ground and the soluble and the membrane fraction were separated by ultra-centrifugation at 100 000 g. Each fraction was incubated with IgG beads and eluted with TEV protease. Plants expressing a TAP tag protein in the wild type background were used as a control. The load, the unbound fraction and the TEV eluate of the soluble (SN100) and membrane (P100) fraction were separated by SDS-PAGE and transferred onto a nitrocellulose membrane for immunoblotting. Toc59 full-length and two smaller fragments were present in the TEV eluate of the membrane fraction together with Toc33 and Toc75. In the TEV eluate of the soluble fraction, only the two smaller fragment of Toc159 were identified. Tic110 was neither in the soluble nor membrane fraction TEV eluate. LHCP, LSU and actin controls demonstrate the separation of the membrane and soluble fractions.

Mass-spectrometry (MS) analysis and filtering of the proteins identified:

The rest of the eluates were loaded on a SDS-PAGE and stained with SYPRO ruby. The lane of each eluate was cut in six different bands which were independently processed. The proteins were digested in gel with trypsin and the tryptic peptides were eluted. The peptides were analyzed by LTQ tandem mass-spectrometry. The obtained spectra were searched with Mascot 2.1.04 (Matrix Science) against "The Arabidopsis Information Resource 8 protein database". The list of the 373 proteins identified in the TAP-Toc 159 eluates of both the soluble and the membrane fraction purifications was filtered as follow: The proteins with less than 4 Peptide Spectrum Matches (PSMs) and which were present in the control eluate were excluded (Please note that Toc159 was identified in the control eluate). The localization of the remaining proteins was searched in the SUBA database (Subcellular Proteomic Database, (Heazlewood et al. 2007)) and assigned their predicted or confirmed location. The SUBA database gives a subcellular location for a given protein, based either on experimental (GFP localization and MS analysis (Hooper et al. 2014)) data or bioinformatics prediction. On the 111 proteins remaining 73 were predicted in the chloroplast, 16 the cytosol and 22 in the vacuole, the mitochondria, the nucleus, the ER, the peroxisome or the plasma membrane. Only the proteins predicted in the chloroplast or the cytosol were kept in the list. The proteins predicted in the chloroplasts were searched for their sub-chloroplastic curated localization (According to the At_Chloro database which predict sub-chloroplastic localization base on MS data and predictions (Ferro et al. 2010). The predictions were as follow: envelope (30), thylakoid (20), stroma (11), un-determined (12). Cytosolic and envelope proteins constitute good candidates for a specific interaction with Toc159, which is exposed to both these cellular compartments. Stromal proteins could also be interesting, as the Toc complex is known to interact with Tic proteins, which in turn are in contact with the stroma. Proteins predicted to the thylakoid were removed of the list. These proteins likely are contaminants or client proteins of Toc159 during import.

Proteins were given an APEX score. The APEX score is a normalized value taking into account the total number of spectra obtained for a given protein, weighted by the total number of spectra obtained for the whole sample and the predicted number of tryptic peptide for the given protein. This value correlates positively with the actual amount of a protein in the sample (Lu et al. 2007). Finally the APEX score of the candidate proteins was normalized by their APEX score in a total leaf extract (Baerenfaller et al. 2008), to lower the weight of abundant proteins which likely are contaminants. The list was ranked according to this ratio (Tab. 1). (The proteins for which APEX score in total leaf extract was not available were put at the end of the list and were ranked according to the PSM obtained). Many of the Toc/Tic proteins (15) were identified in the TAP-Toc159 eluate of the membrane fraction, confirming the results obtained by Western blotting. Interestingly, Tic110 which was not detected by western blot was detected by MS. 5 predicted proteases, 2 predicted kinases and one GAP related protein were identified in the list. Based on the curated localization, the predicted function (and other database annotation) and the phenotype of the mutant of the candidates, three proteins were selected for further experimental characterization. **1)** The protein encoded by the AT4G32250 locus was identified in the chloroplast by a proteomics study and has a predicted protein kinase-like domain. **2)** AT5G01590 has neither predicted sub-chloroplastic localization nor function but it was shown to be co-expressed with FtsH12, a predicted protease also present in the list. The homozygous mutant revealed to have an albino phenotype, reminiscent of *ppi2*. **3)** Emb2004 (At1G10510) is predicted at the envelope of the chloroplast and its closest homologue in *A. thaliana* is a RANGAP protein. The homozygous mutant has an embryo lethal phenotype (Tzafrir et al. 2004) suggesting an important and non-redundant function, like different Toc/Tic homozygous mutants which are embryo-lethal (e.g. Tic75 and Tic110). Moreover, it has a predicted LRR (Leucine Rich Repeat) domain which is often implicated in protein-protein interaction.

Table 1, 1st part: List of the proteins identified in the TAP-Toc159 TEV eluate by mass-spectrometry: The Tap-Toc159 TEV eluate (fig. 9) was analyzed with a LTQ-FTICR mass-spectrometer. The list of identified proteins was filtered for the number of Peptide Spectrum Matches (PSMs) (cut-off = 4), the absence of identification in the control eluate and the predicted sub-cellular localization. (%Cov: coverage of the protein sequence by the identified peptide; Peptides: number of unique peptide identified; APEX: APEX index in this analysis; Norm: APEX in this analysis divided by the APEX in a total leaf extract; Loc.: sub-chloroplastic localization predicted by the At_Chloro Database; MapManBin: functional annotation; blue: Toc/Tic proteins; Grey: Tic 1MD complex; Salmon: Proteases; Yellow: Kinases: Red: GTPase related proteins. Red dot: protein selected for characterization)

Accession	Description	MW (kD)	% Cov.	Peptides	PSMs	APEX	Norm.	Loc.	MapManBin
AT5G56000	HSP81-4	80.14	9.6	6	12	1.07	21.19	Cyt	stress.abiotic.heat
AT2G16640	Toc132	132.270004	6.9	7	9	0.50	15.90	OM	protein.targeting.chloroplast
AT1G13440	GAPC-2	36.91	21.3	6	23	3.77	13.54	Cyt	glycolysis.glyceraldehyde 3-phosphate dehydrogenase
AT5G05000	TOC34	34.70	28.8	8	29	4.57	12.80	OM	protein.targeting.chloroplast
AT5G01590	similar to unnamed protein product	48.87	17	10	12	1.37	10.60	Chl	not assigned.unknown
AT4G31780	MGD1	55.43	9.8	5	6	0.57	10.60	IM	lipid metabolism.glycolipid synthesis.MGDG synthase
AT1G10510	EMB2004 (similar to RAN GAP1)	64.72	25.6	15	39	4.21	9.84	Chl	development.unspecified
AT3G17970	TOC64-III	64.02	25	14	34	3.87	7.51	OM	misc.tetratricopeptide repeat (TPR) unknown function
AT3G25680	similar to unknown protein	63.05	7.2	5	7	0.82	6.18	Chl	not assigned.unknown
AT3G46740	TOC75-III	89.19	34.1	37	137	10.79	5.38	OM	protein.targeting.chloroplast
AT2G36160	40S ribosomal protein S14	16.25	24	4	6	3.07	5.30	Cyt	protein.synthesis.ribosomal protein.eukaryotic.40S subunit.S14
AT4G02510	TOC159	160.82	45.5	90	286	10.95	4.68	OM	protein.targeting.chloroplast
AT1G02280	TOC33	32.92	32.7	8	26	4.63	3.83	OM	protein.targeting.chloroplast
AT3G53560	chloroplast lumen common family protein	38.66	8.5	4	4	0.66	3.53	Chl	not assigned.no ontology
AT2G31610	40S ribosomal protein S3	27.52	21	5	14	3.19	3.53	Cyt	protein.synthesis.ribosomal protein.eukaryotic.40S subunit.S3
AT3G16290	ATPase/ metallopeptidase	99.86	6.4	5	8	0.52	3.53	Chl	protein.degradation.metalloprotease
AT5G02940	similar to phosphotransferase-related	92.13	19.9	19	38	2.68	3.20	Env	not assigned.unknown
AT5G24690	similar to RETICULATA-RELATED 1	56.81	15.5	7	17	1.74	2.73	Chl	not assigned.unknown
AT2G43950	OEP37	38.83	21.3	7	15	2.46	2.65	OM	transport.metabolite transporters at the Env membrane
AT5G22640	Tic100	99.95	7.3	5	8	0.70	2.02	Nuc	not asigned.no ontology
AT5G03910	Arabidopsis ABC transporter homolog 12	69.19	18.1	9	9	0.92	1.99	Env	transport.ABC transporters and multidrug resistance systems
AT4G33520	metal-transporting P-type ATPase 1	99.99	5.9	4	4	0.32	1.77	IM	transport.metal
AT3G04340	EMB2458; ATPase	111.01	6.5	7	7	0.48	1.55	Chl	protein.degradation.metalloprotease
AT5G42480	ARC6	88.26	14	9	15	1.16	1.47	IM	cell.division.Chl
AT5G19620	TOC75-V	79.93	6.4	4	8	0.55	1.41	OM	protein.targeting.chloroplast
AT1G79560	FTSH12 (EMBRYO DEFECTIVE 1047)	115.10	14	13	21	1.28	1.37	Chl	protein.degradation.metalloprotease
AT1G77590	LONG CHAIN ACYL-COA SYNTHETASE 9	76.17	17.5	13	25	2.77	1.30	OM	lipid metabolism.FA synthesis and FA elongation.long chain fatty acid CoA ligase
AT4G23940	FtsH protease, putative	105.54	11.6	10	10	0.65	1.26	IM	protein.degradation.metalloprotease
AT3G14110	FLUORESCENT IN BLUE LIGHT	34.58	18.5	4	7	1.20	1.24	IM-Thy	tetrapyrrole synthesis.regulation
AT3G06510	SFR2, hydrolase	70.77	8.8	4	13	1.16	1.23	OM	lipid metabolism.glycolipid synthesis
AT3G20320	TGD2	41.63	22.3	10	15	2.36	1.20	IM	lipid metabolism.transport
ATCG01130	Tic214 (YCF1)	213.70	19.5	36	36	1.14	1.12	Env	not asigned.no ontology
AT1G20010	tubulin beta-5 chain	50.34	12.5	4	9	1.42	0.99	Cyt	cell.organisation
AT4G37270	Heavy metal ATPase 1	88.19	4	4	4	0.32	0.88	Env	transport.metal
AT5G64580	AAA-type ATPase family protein	96.85	7.4	6	6	0.39	0.88	Chl	protein.degradation.metalloprotease
AT5G50920	Hsp93-V	103.45	27.8	25	70	4.55	0.71	IM	protein.degradation.serine protease
AT1G06950	TIC110	112.12	32.1	33	73	4.67	0.64	IM	protein.targeting.chloroplast
AT2G44640	similar to PIGMENT DEFECTIVE 320	49.83	12.9	4	6	0.82	0.62	Env	not assigned.unknown
AT3G48870	Hsp93-III	105.77	23.2	5	8	0.52	0.57	Str	protein.degradation.serine protease
AT5G16620	Tic40	48.90	15.2	5	8	0.99	0.52	IM	protein.targeting.chloroplast
AT5G49910	cpHSC70-2	76.99	12.3	8	15	1.37	0.52	Str	protein.folding
AT5G12470	similar to unknown proteins vinifera	41.32	10.1	4	11	1.73	0.49	IM	transport.metabolite transporters at the Env membrane
AT1G72370	40S ribosomal protein Sa-1	32.29	15	5	11	2.50	0.47	Cyt	protein.synthesis.ribosomal protein.eukaryotic.40S subunit.SA
AT1G01790	KEFFLUXANTIPORTER 1	64.98	15.2	16	19	1.11	0.37	IM	transport.metabolite transporters at the Env membrane
AT4G25450	A. thaliana non-intrinsic ABC protein 8	77.92	11.7	7	8	0.70	0.36	IM	transport.metabolite transporters at the Env membrane

Toc/Tic proteins	Tic 1 MD complex	Proteases	Kinases	GAP

Table 1, 2nd part: idem as 1st part.

Accession	Description	MW (kD)	% Cov.	Peptides	PSMs	APEX	Norm.	Loc.	MapManBin
AT5G14740	BETA CARBONIC ANHYDRASE 2	36.61	29.7	8	26	3.80	0.36	Str	org.transformation.carbonic anhydrases
AT3G47520	MALATE DEHYDROGENASE	42.40	22.6	11	13	1.78	0.32	Str	org.transformation.other organic acid transformaitons.misc
AT5G46110	APE2; antiporter	44.63	7.4	4	13	2.22	0.32	IM	transport.metabolite transporters at the Env membrane
ATCG00800	ribosomal protein S3	25.18	20.2	5	7	1.69	0.29	Chl	
AT3G01500	CARBONIC ANHYDRASE 1	29.50	32.2	11	41	7.30	0.25	Str	TCA / org.transformation.carbonic anhydrases
AT3G63410	APG1; methyltransferase	37.92	35.5	11	15	2.36	0.23	IM	secondary metabolism.isoprenoids.tocopherol biosynthesis.MPBQ/MSBQ methyltransferase
AT2G38040	acetyl-CoA carboxylase	85.30	4.8	4	5	0.37	0.19	IM	lipid metabolism.FA synthesis and FA elongation.Acetyl CoA Carboxylation
AT5G02500	HSC70-1	71.35	24	6	8	0.82	0.17	Cyt	protein.folding
AT3G09200	60S acidic ribosomal protein P0	34.13	13.9	5	5	1.28	0.14	Cyt	protein.synthesis.ribosomal protein.eukaryotic.60S subunit.P0
AT3G12780	PHOSPHOGLYCERATE KINASE 1	50.11	26.2	10	26	2.88	0.12	Str	PS.calvin cyle.phosphoglycerate kinase
AT3G56940	DICARBOXYLATE DIIRON 1	47.63	12.5	4	4	0.57	0.10	IM	tetrapyrrole synthesis.magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase
AT5G17920	methionine synthase	84.35	8	6	12	0.89	0.06	Cyt	amino acid metabolism.synthesis.aspartate family.methionine
AT2G39730	RUBISCO ACTIVASE	49.10	12.9	10	36	5.27	0.05	Str	PS.calvin cyle.rubisco interacting
AT4G38970	fructose-bisphosphate aldolase, putative	42.98	12.8	5	5	0.60	0.03	Str	PS.calvin cyle.aldolase
AT3G45140	LIPOXYGENASE 2	102.04	7.8	6	6	0.36	0.02	Str	hormone metabolism.jasmonate.synthesis-degradation.lipoxygenase
AT1G56070	Low expression of osmotically responsive genes 1	93.89	5.5	4	5	0.34	0.02	Cyt	protein.synthesis.elongation
				APEX factor not normalized					
ATCG00860	Encodes an unknown protein	269.56	5.8	11	11	0.31		Chl	protein.assembly and cofactor ligation
AT4G32250	protein kinase family protein	68.15	14.2	7	13	1.21		Cyt	protein.postranslational modification.kinase
AT3G13920	DEAD-box ATP-dependent RNA helicase 4	46.71	18.4	7	15	1.98		Cyt	protein.synthesis.initiation; 27.6* RNA DEAD BOX helicase
AT5G19770	tubulin alpha-3	49.66	18.4	6	14	2.29		Cyt	cell.organisation
AT1G65410	A. THALIANA NON-INTRINSIC ABC PROTEIN	37.51	17.1	5	8	1.56		Chl	lipid metabolism.transport
AT2G03510	band 7 family protein	40.60	12.6	5	11	1.73		Cyt	not assigned.no ontology
AT1G33120	60S ribosomal protein L9	22.02	13.9	4	6	1.89		Cyt	protein.synthesis.ribosomal protein.eukaryotic.60S subunit.L9

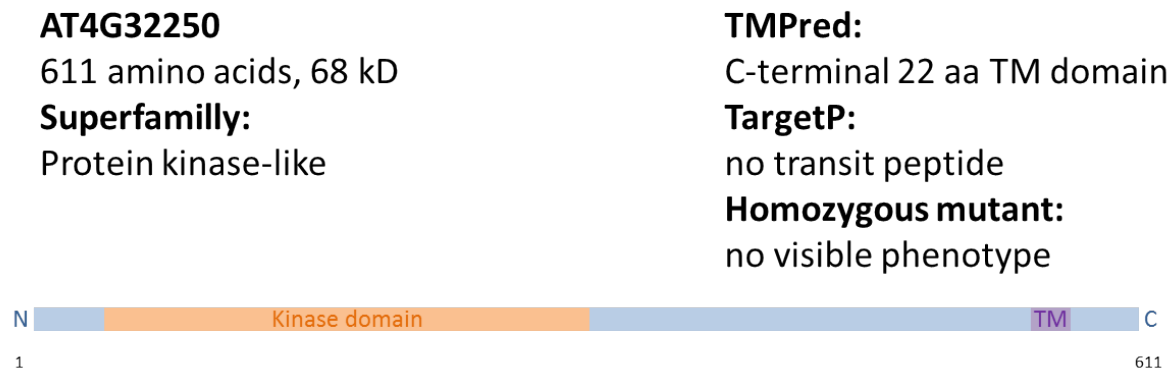
Toc/Tic proteins	Tic 1 MD complex	Proteases	Kinases	GAP

AT4G322250, a predicted protein kinase at the chloroplast envelope:

Bioinformatic data (fig. 10):

AT4G32250 codes for a 611 amino acids proteins of approx. 68 kDa. The prediction of subcellular localization algorithm TargetP (Emanuelsson et al. 2007) predicts no transit peptide in AT4G32250, but different proteomic studies identified it in the plastid proteome (Zybailov et al. 2008). A 22 amino acid transmembrane domain from positions 551 to 572 is predicted by the TMPred algorithm (Hofman & Stoffel 1993), which identifies potential transmembrane domains by comparison with a database of experimentally assessed transmembrane domains. According to the PROSITE database (Sigrist et al. 2013), AT4G32250 has a protein-kinase like domain from amino acid 39 to 306. Its closest protein match in *A. thaliana* is an E3 ligase protein that also contains a kinase domain and is encoded by the gene AT5G13530 (Chen et al. 2013).

a



b

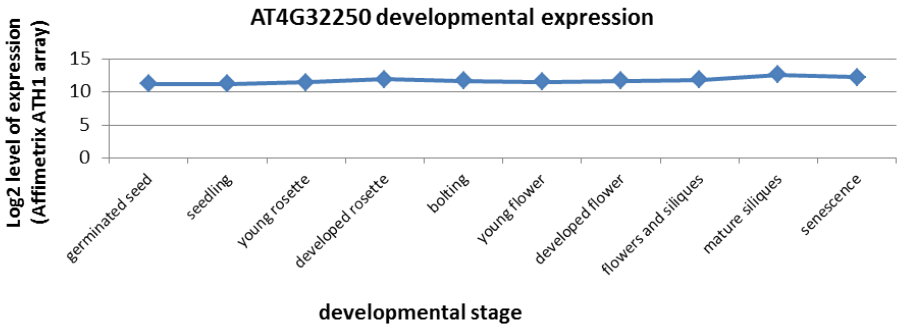


Figure 10: AT4G32250 protein kinase: **a**) Predicted bioinformatics data and linear model of AT4G32250. A kinase domain, a C-terminal transmembrane domain (TM), but no transit peptide was predicted. **b**) Developmental expression profile (Log2 level of expression measured by Affimetrix ATH1 array (GENEVESTIGATOR).

Two insertion mutants from the Salk-Institute collection of T-DNA insertion mutants were obtained. Salk 051823 has a head to tail tandem insertion of two T-DNA elements in the fourth and last exon of AT4G32250 gene. Salk 083378 has a single T-DNA insertion in the third exon of the AT4G32250 gene. Grown on ½ MS + 1% sucrose, these two mutants have no visible phenotype when compared to the wild-type control plant (not shown, Mónica Arias Maldonado, personal communication). The subcellular localization of AT4G32250 was addressed by the *in vitro* import of a radiolabeled AT4G32250 and by YFP fusion protein localization.

***In vitro* import of AT4G322250 (Fig. 11):**

The sequence of AT4G32250 was cloned in a bacterial expression vector, under the control of a T7 promoter. A ^{35}S radiolabeled AT4G32250 was *in vitro* translated (IVT) from this plasmid using an erythrocyte cell-free lysate transcription/translation kit. The IVT product was used for an import/insertion experiment with isolated pea chloroplasts. The import reactions were carried out for 0, 7.5 and 15 minutes. Chloroplasts contained in the 15 minute import reactions were re-isolated and treated with thermolysin in the presence or absence of TX-100. Proteins of the import reaction were precipitated and analyzed by SDS-PAGE and autoradiography (fig. 11). The IVT product band migrates slightly above 66 kDa, consistent with the predicted 68kD size of AT4G322250. The same band is present in the chloroplasts of the import reactions, its intensity increasing with the duration of the import reaction. Under thermolysin treatment, this band disappears while a new approx. 35 kDa band appears. But this band is still present even after the solubilization of the membrane with TX-100 and subsequent protease treatment.

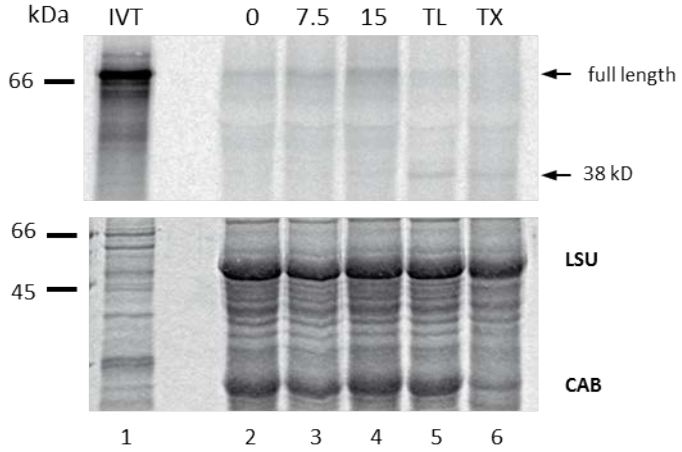


Figure 11: AT4G32250 inserts at the cytosolic side of the chloroplast outer envelope. [^{35}S] Met radiolabeled AT4G32250 was produced using an *in vitro* transcription/translation system (IVT). AT4G32250 IVT was incubated for 0, 7.5 and 15 minutes with isolated pea chloroplasts (lane 2-4). Two 15 minute import reactions were treated with thermolysin (TL, lane 5) or thermolysin and Triton X-100 (TX, lane 6). Proteins of the import reaction were extracted and separated by SDS-PAGE. The dried gel was analyzed by phosphor imaging. AT4G32250 bound to the membrane but was not processed (lane 2-4) and it was sensitive to thermolysin (lane 5). A smaller 38 kDa resistant fragment appeared after the protease treatment. After solubilization of the chloroplast with TX-100 and thermolysin treatment this fragment was still visible, suggesting it rather corresponded to an intrinsically resistant portion of AT4G32250 than a membrane protected one. The lower panel shows the Coomassie stained gel as a loading control. AT4G32250 IVT is shown in lane 1 as a control. CAB = Chlorophyll A/B Binding protein; LSU= Large Subunit of Rubisco.

The absence of a size shift of the band associated with the chloroplasts after the import reaction confirms the absence of a transit peptide predicted by the bioinformatics data. The partial sensitivity to thermolysin indicates that at least half of the AT4G322250 sequence is exposed to the cytosolic side of the chloroplast envelope. As the 35 coda fragment is resistant even after the solubilization of the membrane by a detergent, it is likely that this fragment is not protected by the outer membrane of the chloroplast but rather intrinsically resistant, but neither possibility can be excluded. The Coomassie staining control also shows that the digestion of the chloroplast proteins by the TX-100 detergent plus thermolysin protease is partial but this is probably due to an incomplete digestion by the thermolysin rather than an incomplete solubilization of the membranes. Indeed, the CAB protein, located in the thylakoid membrane appears even more digested than the soluble stromal RBCS.

Transient expression AT4G322250-YFP in tobacco leaves:

The AT4G322250 sequence was cloned in a plant expression vector with an YFP fusion at the C-terminus. The construct was transformed in *A. tumefaciens*. *N. benthamiana* Tobacco leaves were infiltrated with a suspension of an overnight culture of *A. tumefaciens.* bacteria carrying the AT4G322250-YFP construct. The cells in the agro-infiltrated leaf area were observed by confocal microscopy 48h later (fig. 12). The YFP fluorescence signal formed regular thin circles, which coincided with the periphery of the chloroplasts revealed by the chlorophyll auto-fluorescence. Together with the sensitivity to the thermolysin, these data suggest that AT4G322250 is located at the outer membrane of the chloroplast, anchored by the C-terminal predicted TM domain, the N-Terminal kinase domain facing the cytosolic side of the membrane.

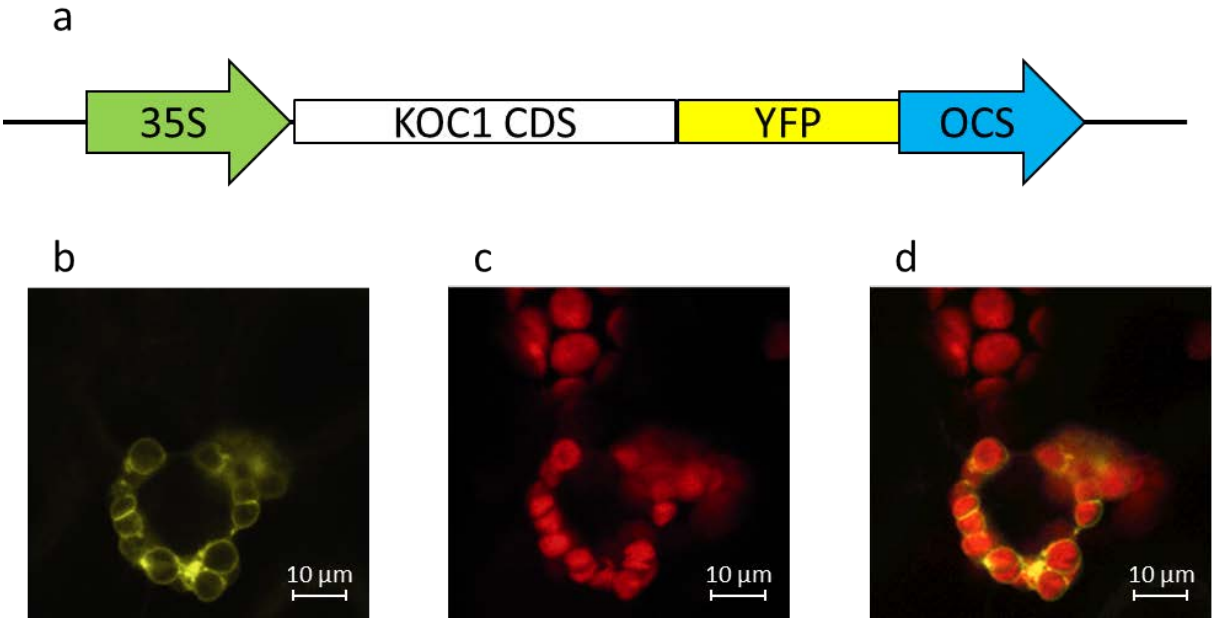


Figure 12: AT4G32250-YFP is localized at the envelope of the chloroplast: a) The AT4G32250 full length sequence was cloned in the plant expression pEARLYGATE-YFP vector to generate a C-terminal fusion with the YFP fluorescent protein. The recombinant gene was under the control of the 35S promoter and had an octopine synthase terminator sequence. *A. tumefaciens* containing the AT4G32250-YFP construct were infiltrated in *N. benthamiana* leaves which were observed by confocal microscopy 3 days after the infiltration. **b)** YFP signal **c)** chlorophyll autofluorescence **d)** merge of YFP and chlorophyll signal.

Co-purification of AT4G322250 with TAP-Toc159 (Fig. 13):

To confirm the interaction between Toc159 and AT4G32250, an antibody was raised against a recombinant AT4G322250 fused to a His Tag. The fusion protein was purified with NiNTA beads and use to raise an antibody. The obtained serum was purified against the antigen used for the immunization. This purified antibody detects a single band in a wild-type plant extract just above 66 kDa, consistent with the 68 kDa predicted size. This band is not visible in a plant extract of an AT4G322250 insertion mutant line (not shown, Mónica Arias Maldonado, personal communication). The same TAP-Toc159 pull-down experiment as was used for the mass-spectrometry analysis was carried out with the following modifications: the control TAP plant extract was not separated into the soluble and membrane fractions before detergent solubilization. The volume of the total TAP fraction was the same as that of each of the soluble and membrane fraction of the TAP-Toc159 plant extract,

to maintain the same concentration of proteins of each type. 50 µg of the load and the flow-through, and an equivalent volume of the first wash of the beads, and 10% of the last wash and of the TEV eluates were loaded for analysis on a SDS-PAGE gel and transferred onto a nitrocellulose membrane and probed with IgG against the recombinant proteins and antibodies against Toc/Tic proteins (fig. 13). The recombinant NTAP and TAP-Toc159 were expressed and present in the TEV eluate (Note that the band between 116 and 97.4 kDa in the load and flow-through lanes in the anti-CBP panel of the NTAP purification corresponds to the residual signal of anti-Tic110 antibody, used in a previous probing of the membrane). The two other members of the Toc core complex, Toc33 and Toc75 were co-purified from the membrane fraction of the TAP-Toc159 extract.

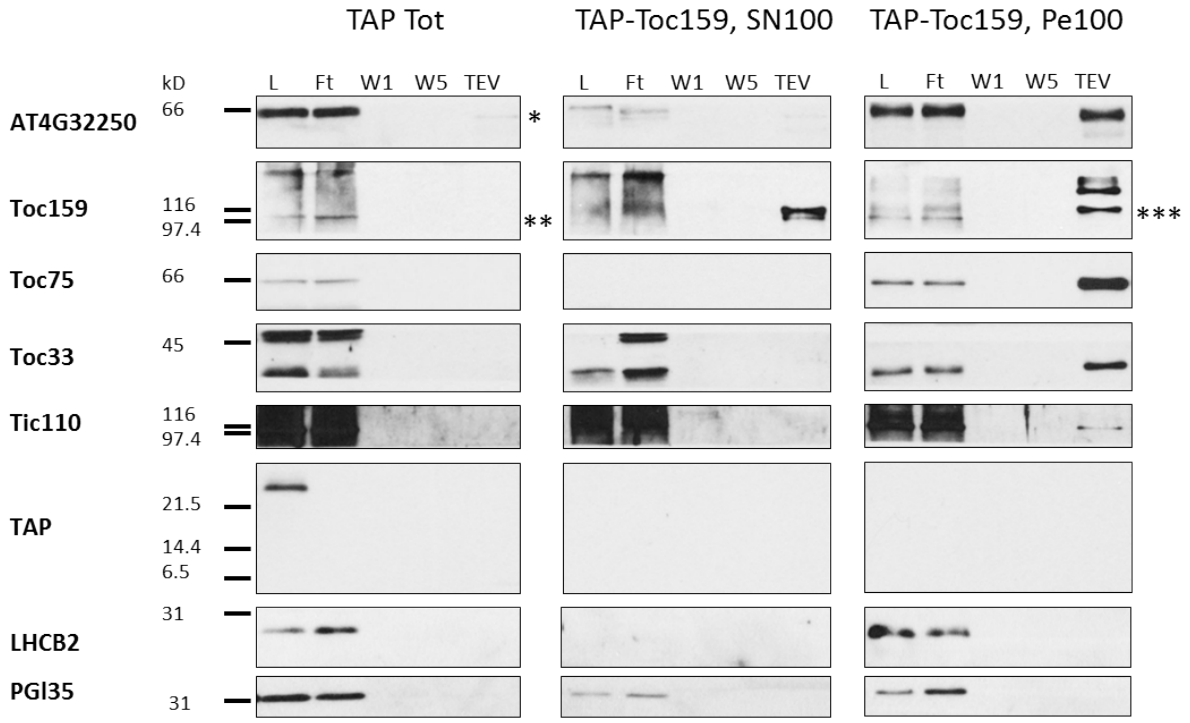


Figure 13 : AT4G32250 interacts in vivo with TAP-Toc159. A TAP-Toc159 plant extract was separated into soluble and membrane fractions by ultracentrifugation at 100 000 g. The membrane fraction (P100) was solubilized with 0.75% TX-100 and incubated with IgG sepharose beads. After extensive washing the beads were eluted by TEV protease. A sample of the load (L), flow through (Ft), 1st wash (W1), last wash (W5), and TEV elution (TEV) were precipitated and separated on a SDS-PAGE gel. Proteins were blotted on a nitrocellulose membrane which was probed with different antibodies. An antibody against the CBP (Calmodulin Binding Protein) portion of the TAP tag was used to detect TAP-Toc159 before and after the TEV cleavage. TAP-Toc159 was present in the protein extract and bound to the IgG beads. The Toc core complex proteins Toc33 and Toc75 co-purified with TAP-Toc159. Tic110 was also co-purified but was much less enriched than the Toc core complex proteins. AT4G32250 was also present in the TEV eluate of the TAP-Toc159 purification, confirming the previous results obtained by mass-spectrometry. *: unspecific band lower than AT4G32250 detected by the anti-AT4G32250 antibody. This band might correspond to the heavy chain of leaked IgG or to a keratin contamination. ** : Residual signal from previous probing with Tic110.***: Band corresponding to the TAP-Toc159 lower band (The ProteinA part of the TAP-tag reacts with any IgG). A total extract of NTAP expressing plants was used as a control. Two membrane proteins, PGL35 and LHCB2 were used as a control for the specificity of the interaction with the IgG beads. Please note that in the Toc33 NTAP panel, the membrane was cut just below the Toc33 bands, partially masking it.

Unlike the first TAP-Toc159 purification experiment shown in fig. 9, Tic110 was detected in the membrane fraction, but much less enriched than Toc75 and Toc33. AT4G322250 was present in the membrane fraction, consistent with its previously described envelope localization. AT4G322250 was

detected in the TAP-Toc159 membrane fraction eluate, but not in the NTAP eluate, confirming its interaction with Toc159. Two faint bands just below that of AT4G322250 were detected in all the eluates, with a similar intensity indicating that these are likely non-specific contaminants. One of these bands might be due to a cross reaction of the anti-AT4G322250 antibodies with leaked IgG heavy chains, which has an Mw of 60 kDa. Neither LHCB2 nor PGL35 are present in any eluates.

The predicted kinase domain of At4G32250, its localization and the confirmed interaction with Toc159 suggest that it might be involved in Toc159 A-domain phosphorylation (fig. 14).

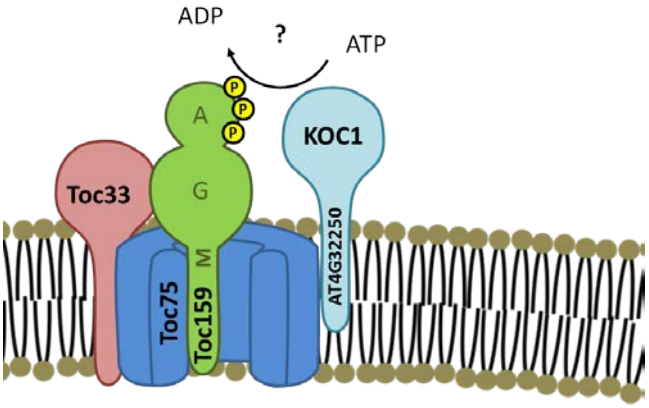


Figure 14 : AT4G32250 as a Toc159 kinase? AT4G32250 is localized at the outer membrane of the chloroplast envelope, probably integrated by a C-terminal transmembrane domain. AT4G32250 co-precipitates with Toc159 and has a predicted kinase function, suggesting that it could phosphorylate the highly phosphorylated A-domain of Toc159.

AT5G01590-Tic56:

AT5G01590 encodes Tic56, a component of the 1MD Tic complex, identified by(Shingo Kikuchi et al. 2013). It is a 527 amino acids and 62 kDa protein, with a 48 aa predicted transit peptide (Target P) (fig. 15, a). No known protein domain is reported based on protein sequence comparison. The expression of Tic56 is fairly constant till the bolting stage and then declines during senescence (fig. 15, b). The T-DNA insertion mutant tic56-1 (SAIL_810_G07) exhibits an albino phenotype when grown ½ MS + 1% sucrose (fig. 16, a). A cross section of fixed albino *tic56-1* plants was observed by light microscopy. Epidermal cells appeared bigger than those of the wt control and the mesophyll cells are less abundant and disorganized, the chloroplasts were barely visible (fig. 16, b). The ultrastructure of the *tic56-1* chloroplasts was observed in more detail by TEM microscopy, revealing their small size, the scarcity of thylakoid membranes and the absence of thylakoid grana (fig. 16, c) (Köhler et al. 2015).

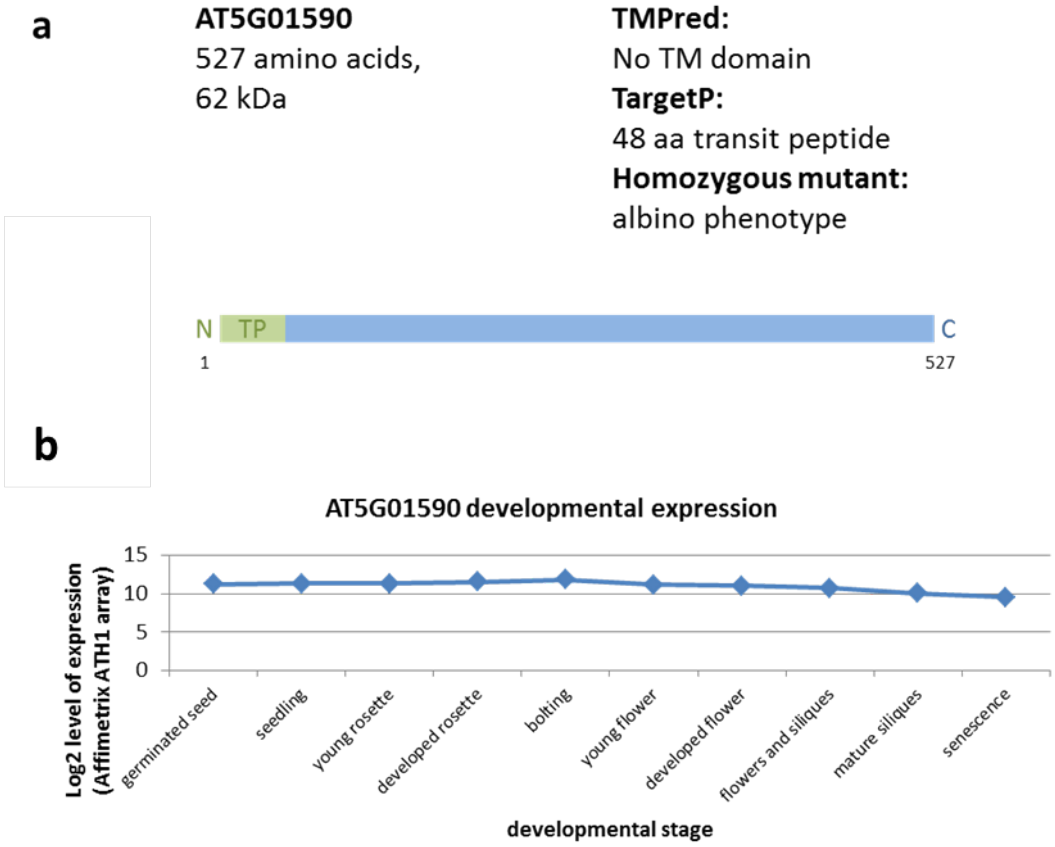


Figure 15: AT5G01590-Tic56: a) Bioinformatic predictions and linear model of the protein sequence of AT5G01590-Tic56. A 48 aa transit peptide was predicted but no transit peptide. **b) b**) Developmental expression profile of AT5G01590-Tic56 (Log2 level of expression measured by Affimetrix ATH1 array (GENEVESTIGATOR).

To confirm the interaction of Tic56 with Toc159, a TAP-Toc159 purification similar to that for the mass-spectrometry analysis was done and the load, flow through, 1st wash, last wash and TEV elution fraction of the purification were analyzed by SDS-PAGE and Western blotting. As in the two TAP-Toc159 purifications already described, the members of the Toc core complex, Toc75 and Toc33 were co-isolated together with TAP-Toc159. Interestingly, Tic110 which was already identified in one of

the two previous TAP-Toc159 purifications was identified as well as Tic40 in the TAP-Toc159 eluate. Tic56 was detected in the TAP-Toc159 eluate, confirming the mass spectrometric analysis of the TEV eluate of the first TAP-Toc159 described previously (The production and characterization of the anti-Tic56 antibody were described in Köhler et al. 2015)). A second member of the Tic 1MD complex, Tic20 was also found in the TAP-Toc159 TEV eluate (fig. 17 and fig. 18).

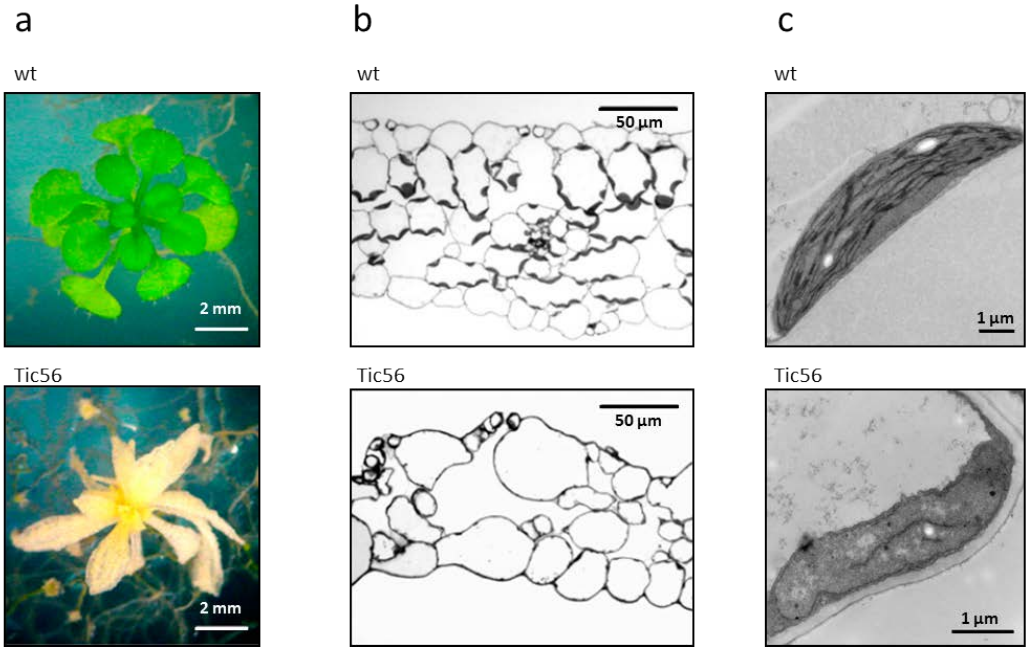


Figure 16 : The homozygous mutant of AT5G01590 (Tic56) is albino and has poorly developed chloroplasts. a) 8 weeks old a *tic56-1* homozygous mutant grown on ½ MS 0.8% sucrose showed an albino phenotype. B) Cross section of a *tic56-1* mutant leaf observed by light microscopy revealed bigger epidermal cell compared to a wild type leaf and the strong reduction of the mesophyll tissues. Chloroplasts were not visible. C) An electron micrograph revealed of a cross section of a *tic56-1* mutant leaf showed the reduced size of the chloroplast, the scarcity of thylakoid membranes and the absence of thylakoid grana. (Adapted from Koehler *et al* (2015))

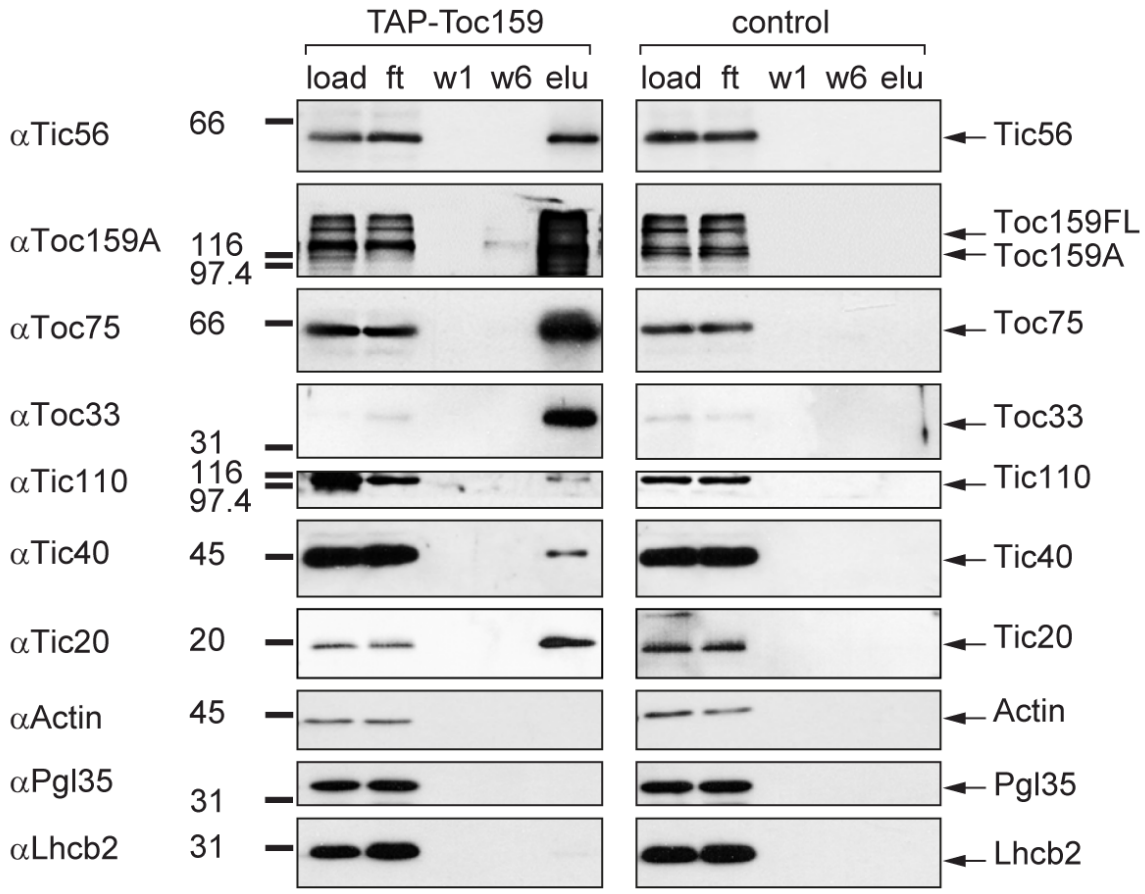


Figure 17 : Tic56 interacts in vivo with TAP-Toc159. A TAP-Toc159 plant extract was separated into soluble and membrane fractions by centrifugation at 100 000 g. The membrane fraction (P100) was solubilized with 0.75% TX-100 and incubated with IgG sepharose beads. After extensive washing, beads were eluted by TEV protease. A sample of the load (L), flow through (Ft) , 1st wash (W1), last wash (W5), and TEV elution (TEV) were precipitated and loaded on a SDS-PAGE gel. Proteins were blotted onto a nitrocellulose membrane which was probed with different antibodies as indicated. Toc159 full length (Toc159FL) and the Toc159 A-domain (Toc159FL) were present in the TAP-Toc159 TEV eluate. The Toc core complex proteins Toc33 and Toc75 co-purified with TAP-Toc159. Tic110 was also co-purified but was much less enriched than the Toc core complex proteins. Tic56 as well as Tic20 were also present in the TEV eluate of the TAP-Toc159 purification, confirming the previous results obtained by mass-spectrometry. A total extract of NTAP expressing plants was used as a control. PGL35 and LHCB2 and Actin were used as a control for the specificity of the interaction with the IgG beads.

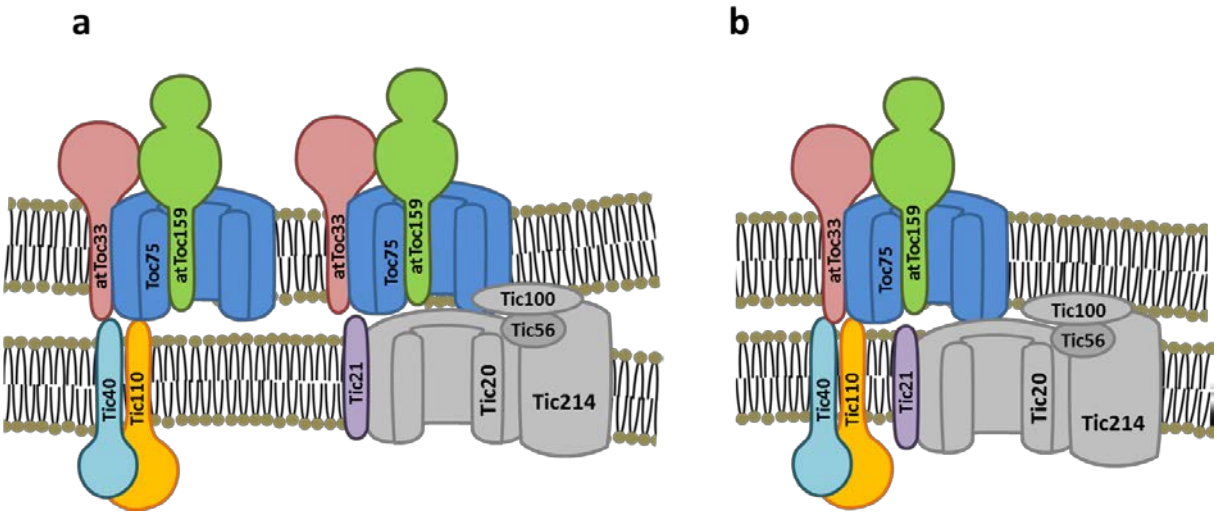


Figure 18 : Tic complex organization: Tic complex organization: Two alternative models can account for the interaction of Toc159 with the components of the Tic 1 MD complex in the absence of externally added precursor protein. **a)** Two different Toc/Tic supercomplexes exist. One is constituted of the Toc core complex and Tic110 and Tic40, the second one formed of the Toc core complex and the 1 MD Tic complex. **b)** Tic110 and Tic40 and the 1MD Tic complex both interact simultaneously with the Toc core complex, to form a Toc/Tic supercomplex.

Emb2004-AT1G10510:

Emb2004 was identified by mass spectrometric analysis of the purified TAP-Toc159 complex. Based on the APEX values Emb2004 was one of the best candidates for a specific component of the TAP-Toc159 complex. In the following, I characterize the protein, attempt to confirm its association with TAP-Toc159 complex and to identify new interacting partners of Emb2004 using both TAP-based isolation and yeast two-hybrid screening.

Emb2004, a LRR protein (fig. 19 and fig. 20):

Emb2004 is a 605 amino acid protein with a predicted molecular weight of 64 kDa. The closest *A. thaliana* homologue is a RanGAP1 (Ran GTPase Activating Protein)(AT3G63130)(Pay et al. 2002). Emb2004 has a predicted transit peptide of 60 amino acids (ChloroP (Emanuelsson et al. 1999)) and a N-terminal transmembrane domain of 19 amino acids. Emb2004 has a conserved domain belonging to the RNI-like (Ribonuclease Inhibitor) superfamily and of the 28-residue LRR (leucine Rich Repeat) family (Superfam(Wilson et al. 2009)). The PROSITE database identifies 9 such conserved motives along the Emb2004 sequence (fig. 19: a, c). LRR are 20-30 amino acids motives enriched in leucine or other hydrophobic amino acids. The secondary structure is composed of tandem alpha helices and beta-sheet arranged in a horseshoes shape at the tertiary structure level (Forsthoefel et al. 2005). The tertiary structure of Emb2004 was modelled using the automated procedure provided by the SWISSMODEL server (Arnold et al. 2006). First, a template is identified based on the sequence similarity. Second, the query protein is fitted against a crystallographic structure of the template. The Emb2004 model obtained displayed typical LRR secondary and tertiary structures (fig. 19, b).

Emb2004 was identified in a screen for embryo lethal mutants (Tzafrir et al. 2004). Indeed, homozygous *emb2004* mutants are arrested at the globular stage of the embryo development. The expression of Emb2004 is high and stable through the development of the plant, suggesting that Emb2004 is not only needed at the embryo development stage (fig. 19, d). The protein sequence of Emb2004 was searched against the NCBI protein sequences collection with the BLAST algorithm (Johnson et al. 2008), and the best match of 4 different families were used for an alignment using the CLUSTALW2 algorithm (Larkin et al. 2007) (fig. 20). The coverage and similarity with Emb2004 of the sequences identified by the BLAST algorithm were as follow: *Tarenia hasslerianna* (Cleomaceae): NLRC3 isoform: 91%, 80%; *Prunus mume* (Rosaceae): NLRC3 isoform: 93%, 75%; *Theobroma cacao* (Malvacea): RNAI-like protein: 99%, 74%; *Populus euphratica* (Saliaceae) NLRC3 protein: 99%, 71%. The sequences are highly conserved in the part of the Emb2004 sequence corresponding to the predicted LRR motives. The N-terminal part is less conserved and corresponds to the predicted region of the transit peptide and the transmembrane domain in the Emb2004 sequence. This highly conserved sequence across species is consistent with the embryo lethal phenotype observed in *A. thaliana* and support a role for the LRR structures in the function Emb2004. Emb2004 was localized in the chloroplast by different proteomic studies (Froehlich et al.; Rolland et al. 2003; Ferro et al. 2003). To confirm and refine the localization of Emb2004, *in vitro* import experiments, protease protection assays and YFP localization were used.

.

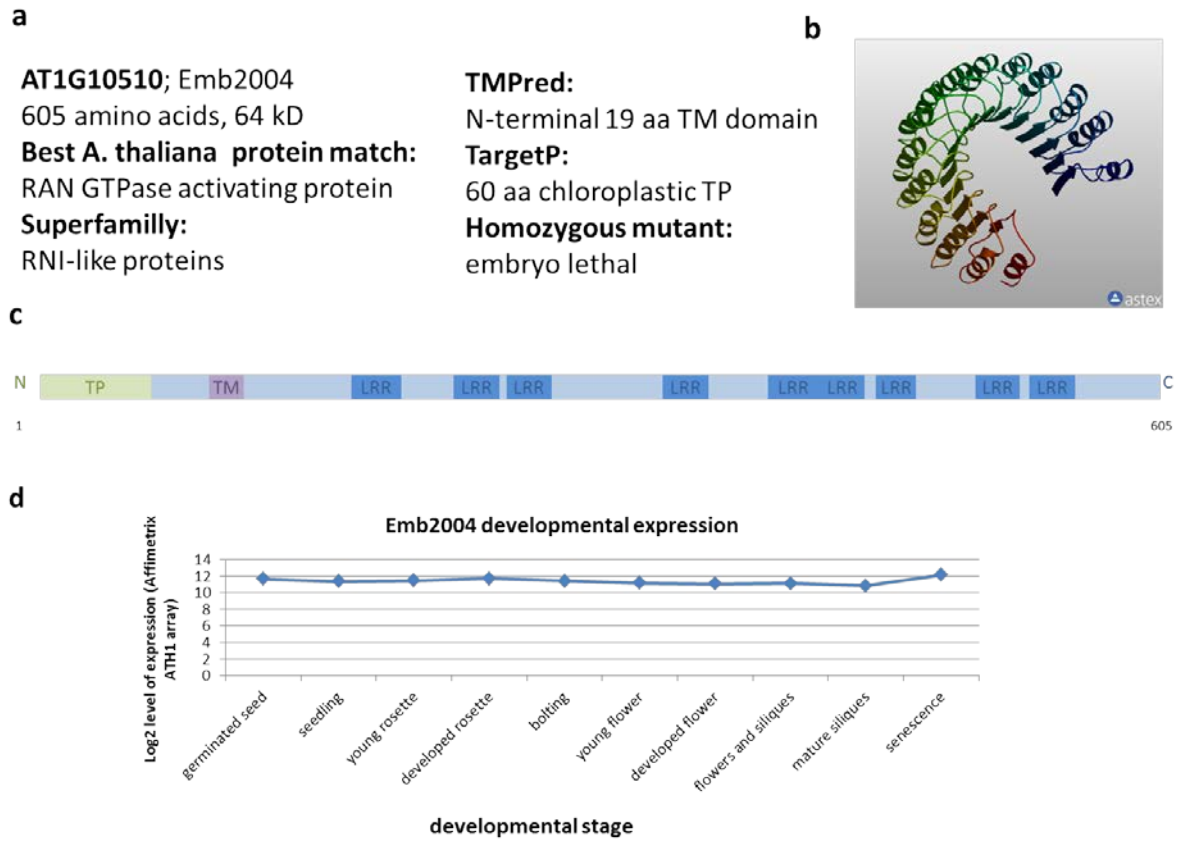


Figure 19: Emb2004 is a LRR protein: a) Emb2004 (AT1G10510) encodes a 605 amino acids, 64 kDa protein: The closest homologue of Emb2004 in the *A. thaliana* genome is the RAN GTPase activating protein encoded by the AT3G63130 locus. The TMpred algorithm predicts a single 19 amino acid transmembrane domain at the N-terminus of the protein sequence. The TargetP algorithm predicts a 60 amino acids long chloroplast transit peptide. The Superfam database reports 9 Leucine Rich Repeat along the Emb2004 sequence. Emb2004 belongs to the Ribonuclease inhibitor-like (RNI-like) superfamily. **b)** A tertiary structure model of Emb2004 was generated by structure homology modeling, based on the crystallographic structure of the Ribonuclease Inhibitor protein (SWISSMODEL). Emb2004 structure displays a typical LRR protein structure, consisting of beta sheet-alpha helix tandem repeat secondary structure arranged in a horseshoe-like tertiary structure. **c)** Linear model of the Emb2004 protein sequence highlighting the predicted transit peptide (TP, green box), the predicted transmembrane domain (TM, violet box) and the LRRs motives (LRR, dark blue boxes) **d)** developmental expression profile of *Emb2004*

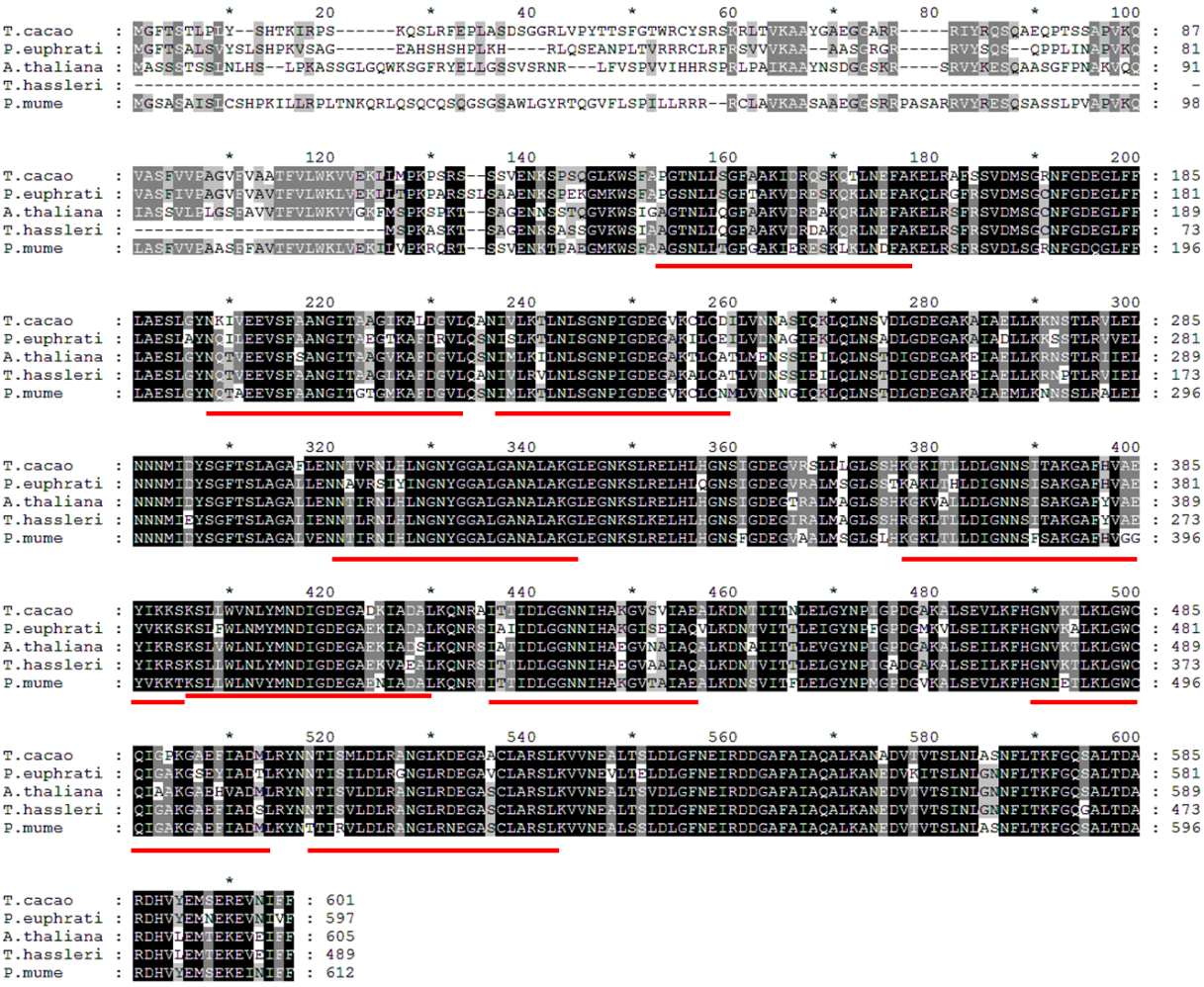


Figure 20: Emb2004 is a well conserved protein across plant species: The Emb2004 protein sequence was searched against the NCBI proteins sequence database using the BLAST algorithm and the first match of different families were selected and aligned using the CLUSTALW2 algorithm. Name of the proteins, % identity and % coverage of the BLAST alignment are as follows: *Tarenia hasslerianna* (Cleomaceae): NLRC3 isoform: 91%, 80%; *Prunus mume* (Rosaceae): NLRC3 isoform: 93%, 75%; *Theobroma cacao* (Malvacea): RNAI-like protein: 99%, 74%; *Populus euphratica* (Saliaceae) NLRC3 protein: 99%, 71%. The four degrees of shading indicate the degree of conservation among the four sequences. The red lines indicate the conserved LRR motives predicted by the PROSITE database.

In vitro import of Emb2004 and protease protection assay (fig. 21):

A ^{35}S Met radiolabeled full length Emb2004 was produced in a reticulocyte cell-free in vitro translation system (IVT). The ^{35}S Emb2004 was used as a substrate in an in vitro import/insertion experiment, using isolated *A. thaliana* chloroplasts. 5 µl of the Emb20040 IVT were incubated with 25 µg equivalent chlorophyll of isolated chloroplasts for 0, 5, 10 and 15 minutes. The reactions were stopped by a 10-fold dilution with cold isotonic buffer, and washed to eliminate unspecific binding. Re-isolated chloroplasts from a 15 minute import reaction were treated with thermolysin. Protein contained in the reactions was separated by SDS-PAGE, 1 µl of the Emb2004 IVT being used as a control. The resulting gels was dried and exposed to a phosphor screen which was analyzed using a Phosphor imager scanner (fig. 21, a). One major band at 66 kDa was present in the Emb2004 IVT consistent with the expected mass of 65 kDa of full length Emb2004. At 0 min of incubation time with

the chloroplast, the full length Emb2004 was bound to the chloroplast. After 5 min a second lower band appeared at ca. 57 kDa, consistent with the 58 kDa predicted mass of mature Emb2004. In the thermolysin treated reaction, the full length Emb2004 band was digested whereas the second lower band was resistant. Taken together, these data show that the radiolabeled Emb2004 is imported in the chloroplasts and processed, confirming the bioinformatics prediction. To further characterize the localization of Emb2004, an additional protease protection assay was done, using both thermolysin and trypsin. An import reaction similar to the one already described in previous paragraph, was carried out. Chloroplasts re-isolated from the 15 min import reactions were treated with two concentrations of Thermolysin (5 and 50 µg ml) and two concentrations of trypsin (5 and 50 µg/ml) followed by incubation on ice for 30 min (fig. 21, b). The full length Emb2004 band was digested by both proteases at the two concentrations tested, confirming the activity of the thermolysin and trypsin. The mature band of Emb2004 was resistant to both proteases at the two concentrations tested. Thermolysin is known to digest only proteins exposed at the cytosolic surface of the chloroplast envelope (Froehlich 2011).

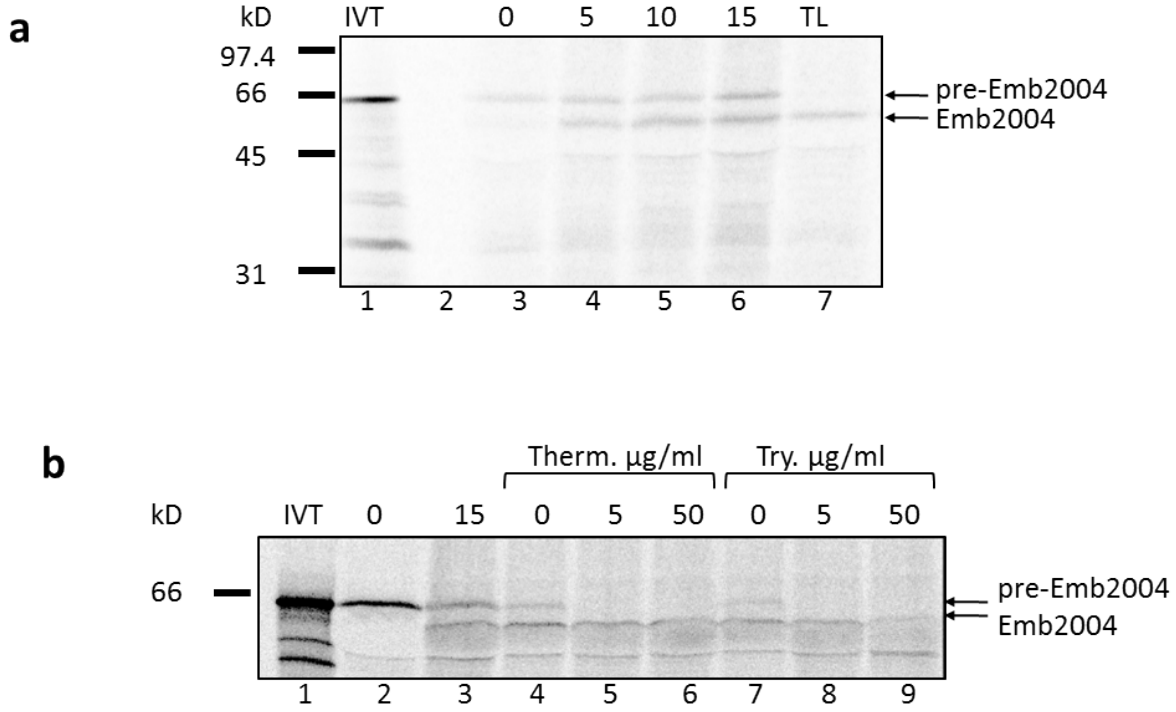


Figure 21: Emb2004 IVT is imported in the chloroplast: a) Radiolabeled [^{35}S]-Met Emb2004 was in vitro translated using a rabbit reticulocyte extract *in vitro* transcription/translation system (lane 1). The import reactions were done by incubating isolated *A. thaliana* chloroplasts with the radiolabeled Emb2004 for 0, 5, 10 and 15 minutes (lanes 3-6). One 15 min import reaction was digested with thermolysin (15 min + TL). The import reactions Proteins were separated on a SDS-PAGE gel which was dried and analyzed by phosphor imaging. **b)** 15 minutes import reactions as described in **a** were incubated with 0, 5 and 50 µg/ml of thermolysin (lanes 5, 6) or trypsin (lanes 8, 9). 0 and 15 min import reaction are shown as control (lane 2-3). Arrows indicate the size of the pre-pre-Emb2004 and the mature Emb2004.

All the proteins inside the outer membrane of the chloroplast envelope are protected. Trypsin is known to digest proteins exposed at the cytosolic side of the chloroplast envelope, but it also penetrates the outer membrane of the envelope and digests the proteins exposed to the intermembrane space. No control for the actual digestion of an intermembrane space marker protein is shown, but the trypsin assay used for this study is commonly used to investigate the topology of the chloroplastic proteins (Froehlich 2011). The results therefore suggest that the majority of EMB2004 is exposed to chloroplast stroma.

Emb2004-YFP localization (fig: 22):

The full length coding sequence of Emb2004 was cloned in the pEARLYGATE-YFP plant expression vector under the control of the 35S promoter and with the Octopine Synthase terminator in 3', to obtain a C-terminal fusion protein of YFP with Emb2004. The EMB2004-YFP construct was transiently transformed in Col2 wt *A. thaliana* protoplasts, using a polyethylene glycol (PEG) transformation method. The protoplasts were observed 24h after the transformation to monitor the expression of the Emb2004-YFP. When merging two channels, the Emb2004-YFP localized at the periphery of the red chlorophyll fluorescence signal of the chloroplasts. No signal was found inside the chloroplast or in the cytosol. This suggests that the Emb2004-YFP fusion protein is localized at the envelope of the chloroplast.

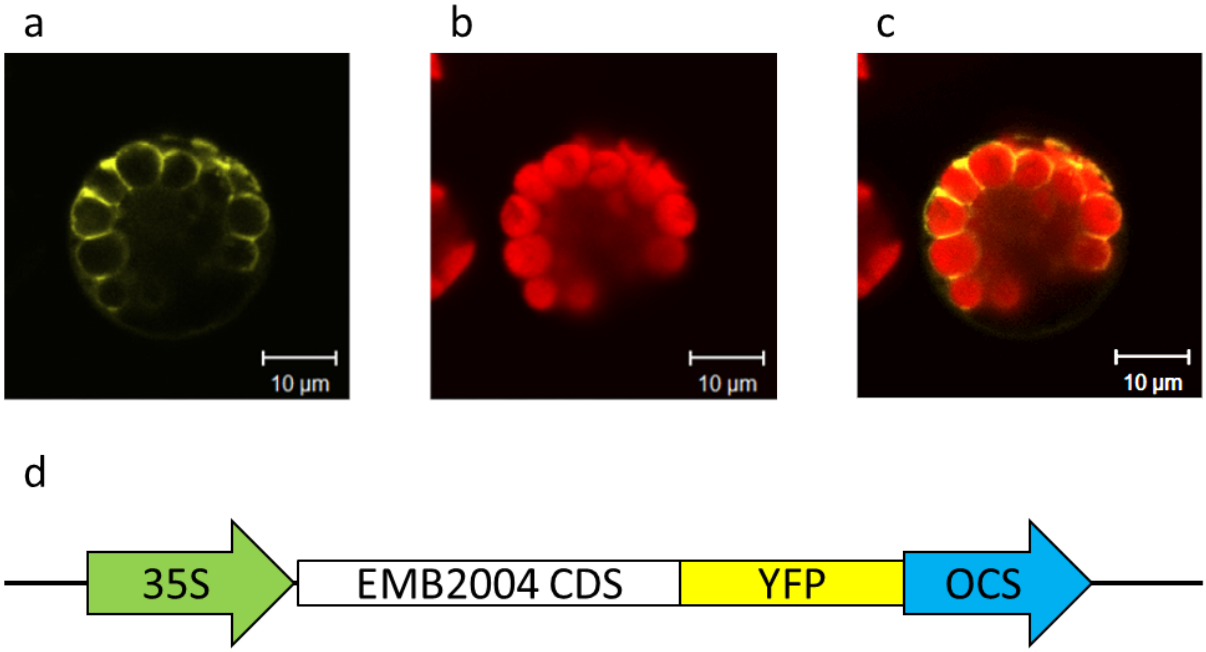


Figure 22 : Transient expression of an EMB2004-YFP fusion protein in *A. thaliana* protoplasts. A pEarlyGate construct coding for EMB2004 fused at its C-terminal to YFP was PEG transformed into isolated *A. thaliana* protoplasts. The fluorescence of the expressed protein (a) and the autofluorescence of the chlorophyll (b) were analyzed by scanning confocal microscopy. A merge (c) of both channels is also displayed. A schematic representation of the construct used for transformation is shown in (d): CaMV 35S promoter (35S); EMB2004 coding sequence (EMB2004 CDS); yellow fluorescent protein sequence (YFP); octopine synthase terminator (OCS).

Sub-chloroplastic localization of Emb2004 (Fig. 23):

Emb2004 has a predicted 17 amino acids hydrophobic transmembrane domain from amino acid 94 to 110. The objective of this experiment was to experimentally determine whether Emb2004 is an integral membrane protein. Two parallel approaches were used: one relying on *in vitro* import of radiolabeled synthetic ^{35}S-Emb2004 into wild type chloroplasts and the other relying on chloroplasts isolated from plants stably expressing an Emb2004-TAP fusion protein (see fig. 28). Chloroplasts were broken by osmotic shock using cold 2 mM EDTA. The membrane and stromal fractions were separated by centrifugation at 15 000 g. The membranes were re-suspended in cold 0.1 M Na_2C0_3 pH 11.5 and ultracentrifuged at 40 000 g. Alkaline Na_2C0_3 treatment of the membrane detaches the proteins which are not integral to the membrane. The ^{35}S-Emb2004 contained in the IVT was imported and processed to the

mature size. The mature band (and the pre-protein band too) was found only in the membrane fraction after the separation of stroma and membranes (fig. 23, lane 8 -9). The mature ^{35}S-Emb2004 was resistant to the alkaline treatment of the chloroplast membrane fraction suggesting that it is an integral membrane protein (fig. 23 , lane 12-13). Emb2004-TAP shows the same association with the membrane fractions. This is in agreement with the predicted N-terminal transmembrane domain, and it can be postulated that Emb2004 is anchored in the membrane by this unique sequence. In the lane corresponding to the stromal fraction (Figure 23, lane 6-7), the RBCL band on the amido-black control appeared abnormally weak suggesting that not all the stromal proteins had been precipitated during the $CHCl_3$/MeOH precipitation. The higher volume i.e. dilution of the stromal fraction compared to that of the membrane might be the cause. However, the amount of the pre-protein and the mature protein detected in the membrane fraction is similar to the one in the total import reaction, suggesting that it is unlikely that the pre-Emb2004 or mature Emb2004 were present at high levels in the stromal fraction. The pre-protein band was also resistant to the alkaline extraction. The pre-protein bound to the chloroplasts after an import experiment may represent an early import intermediate (fig. 3), already engaged in the Toc75 channel and in contact with some of the Tic proteins. Alkaline extraction normally releases membrane-associated proteins which are not inserted in the membrane through a hydrophobic domain, but import intermediates were shown to be resistant to alkaline treatment (Waegemann & Soll 1991). To summarize the in vivo and in vitro data obtained about the localization of Emb2004, the following topology of Emb2004 is proposed: Emb2004 is inserted at the internal membrane of the chloroplast envelope through a single hydrophobic N-terminal transmembrane domain, the C-terminal part of the protein facing the stromal side (fig. 24). In support to the Emb2004-YFP localization presented above, Emb2004 was identified in the inner membrane of the envelope by proteomics (Rolland et al. 2003).

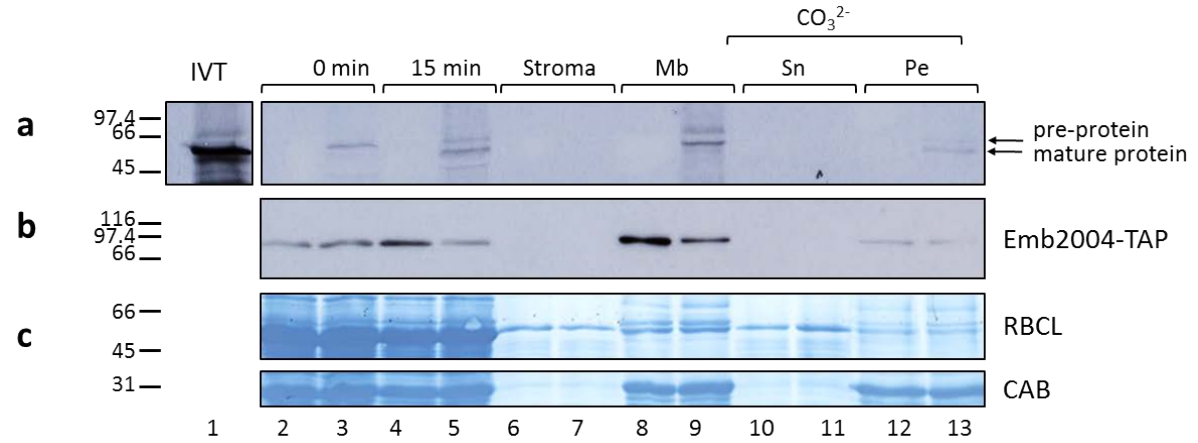


Figure 23: Imported radiolabeled ^{35}S-Emb2004 and Emb2004-TAP are both resistant to alkaline extraction: radiolabeled Emb2004 was imported in isolated Emb400-TAP chloroplast (lane 3 and 5; control: lane 2 and 4). The chloroplasts were broken with osmotical shock and the stromal and the membrane fraction were separated by centrifugation (lane 6-9). The membrane fraction was extracted with CO_3^{2-} pH 11.5 and the soluble and membrane fractions were separated by centrifugation (lane 10-13). The proteins were extracted, separated by SDS-PAGE electrophoresis and transferred on a nitrocellulose membrane. The membrane was exposed to an autoradiography film (a) and probed with IgG (b). Amido-black coloration of the membrane is shown as a loading control (c).

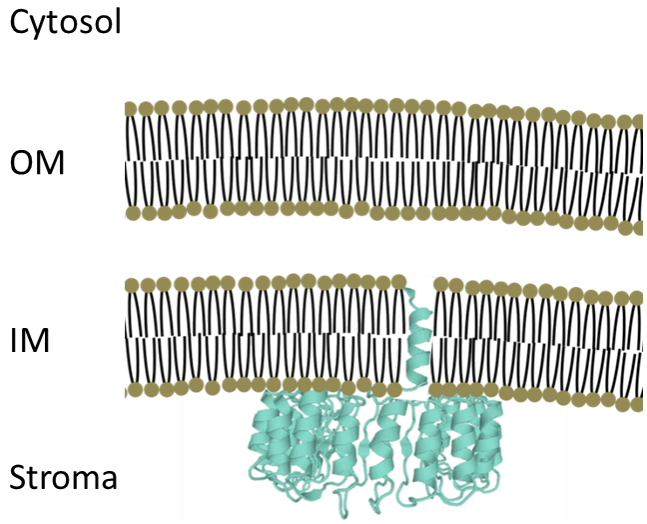


Figure 24: Proposed model for Emb2004 localization: Emb2004 in an integral membrane protein inserted in the inner envelope membrane of the chloroplast and faces the stroma.

Antibody production (Fig. 25):

The Emb2004 full length sequence fused C-terminally to the Maltose Binding Protein (MBP) protein (Emb2004-MBP) was over expressed in *E. coli* using the pETM40 vector system. A solubility test showed that the bulk of the recombinant protein was insoluble. The insoluble inclusion bodies containing Emb2004-MBP were washed several times with 4M urea, re-suspended in PBS and dried. The dried inclusion bodies were sent to Eurogentech for the production of a rabbit polyclonal antibody. To prepare affinity purified antibodies from the immune serum, Emb2004 –MBP was solubilized from inclusion bodies using 8M urea and fixed on Affigel beads. The serum was incubated with Emb2004-MBP resin. IgG bound to the affinity column was eluted with an acidic glycine buffer and concentrated. As the homozygous *emb2004* mutant is embryo lethal, no *emb2004* could be prepared to assess the specificity of the antibody by Western blotting. Instead, ^{35}S-Emb2004 was produced using the reticulocyte cell free in vitro transcription/translation system (IVT). The ^{35}S-Emb2004 was separated by SDS-PAGE and transferred to a nitrocellulose membrane (fig. 25, lane 3-4 and 7-8). An identical volume of reticulocyte that was not programmed to synthesize ^{35}S-Emb2004 was used as a negative control (fig. 25, lane 1-2 and 5-6). The membrane was exposed to an X-ray sensitive film to verify synthesis of ^{35}S-Emb2004 (fig. 25, a). One major band at 66 kDa was obtained in the Emb2004 IVT lane on the autoradiography, consistent with the 65 kDa predicted mass of full length Emb2004. Subsequently, a 1/500 dilution of the anti-Emb2004 purified antibody was used to probe the membrane (fig. 25, b). The anti-Emb2004 antibody detected a specific band of 66 kDa in the lane containing ^{35}S-Emb2004. A cross-reacting band just above 66 kDa was also present both in ^{35}S-Emb2004 and the control lane. The amido-black staining of the membrane accounts for the equal loading of the reticulocyte lysate in the control and Emb2004 IVT lanes. Affinity purified anti-Emb2004 was also used to analyze whole-plant and chloroplast extract.

However, numerous non-specific bands were observed precluding the unambiguous assignment of a band to Emb2004 (not shown).

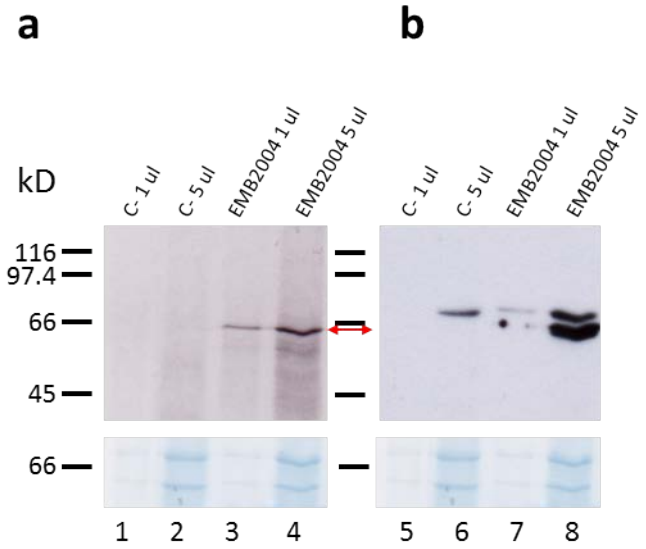


Figure 25 : Anti-Emb2004 purified antibody recognizes in vitro translated Emb2004: : ^{35}S-Met labeled Emb2004 was in vitro translated as described in fig. 21. 1 and 5 µl of the Emb2004 radiolabeled product (lane3-4, 7-8) were loaded on a SDS-PAGE gel along with 1 and 5 µl of non-programmed *in vitro* translation extract as a negative control (lane 1-2, and 5-6). After separation, proteins were transferred to a nitrocellulose membrane. The membrane was analyzed by autoradiography (a) and then probed with anti-emb2004 purified antibody (b). Amido-black staining of the membrane is shown as a loading control. The two-headed arrow indicates the position of Emb2004.

Emb2004 was not detected in a TAP-Toc159 eluate using the anti-Emb2004 antibody (fig. 25):

To obtain additional evidence for the association of Emb2004 with the chloroplast protein import machinery, Western blots containing the purified TAP-Toc159 complex were probed with affinity purified anti-Emb2004. First, the membrane shown in fig. 9 was probed with the anti-Emb2004 antibody (fig. 26). No signals of the correct mass of the mature protein predicted size of about 58 kDa was detected in the eluate. Therefore, a new TAP-Toc159 purification experiment was carried out but gave the same negative result (data not shown).

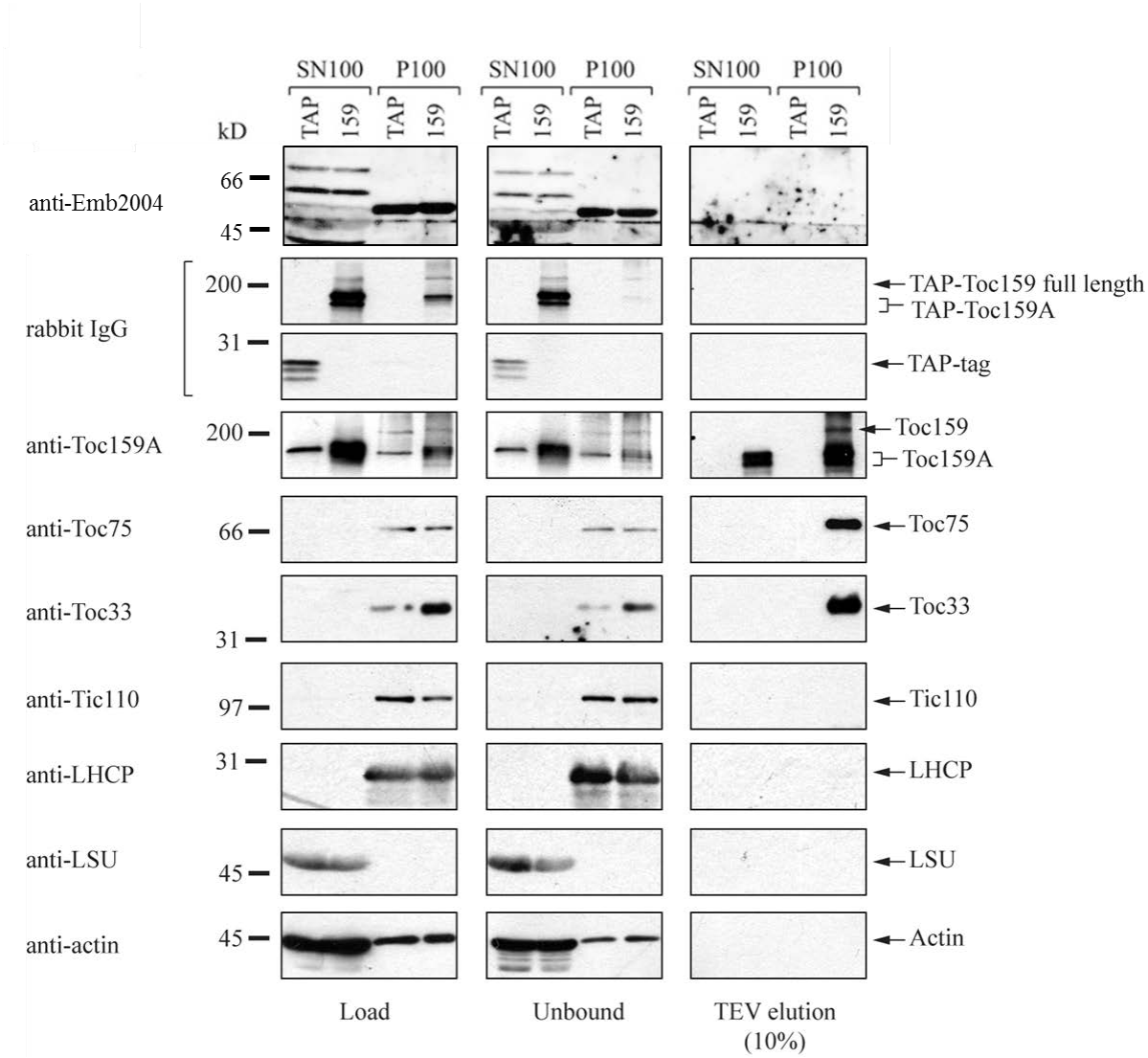


Figure 26 : EMB2004 is not detected in a TAP-Toc159 eluate using anti-emb2004 for Western blotting: The same blot as shown in fig. 9 was probed with the anti-Emb2004 antibody. No signal was detected between 45 and 66 kDa, the expected size of processed Emb2004 (approx. 57 kDa). The absence of an Emb2004 signal in the TEV could be due to the actual absence of Emb2004, but also to the poor sensitivity of the anti-Emb2004 antibodies.

Yeast two hybrid assay (Fig. 26):

In our efforts to obtain evidence for an interaction between Toc159 and Emb2004 we therefore decided to carry out a Yeast Two Hybrid screen. To test the ability of Emb2004 to bind Toc159 or one of the other Toc core complex component, pairwise yeast two hybrid assays were done. The two-hybrid system used was based on the complementation of the GAL4 transcription factor. Gal4 has one domain that binds to the DNA, and one activating domain which recruits the RNA polymerase. Each protein to be tested is fused to either the activating or the binding domain and expressed in the same yeast cells. Positive interaction between the two tested proteins joins the two separated Gal4 domain into a functional transcription factor which activates reporter genes.

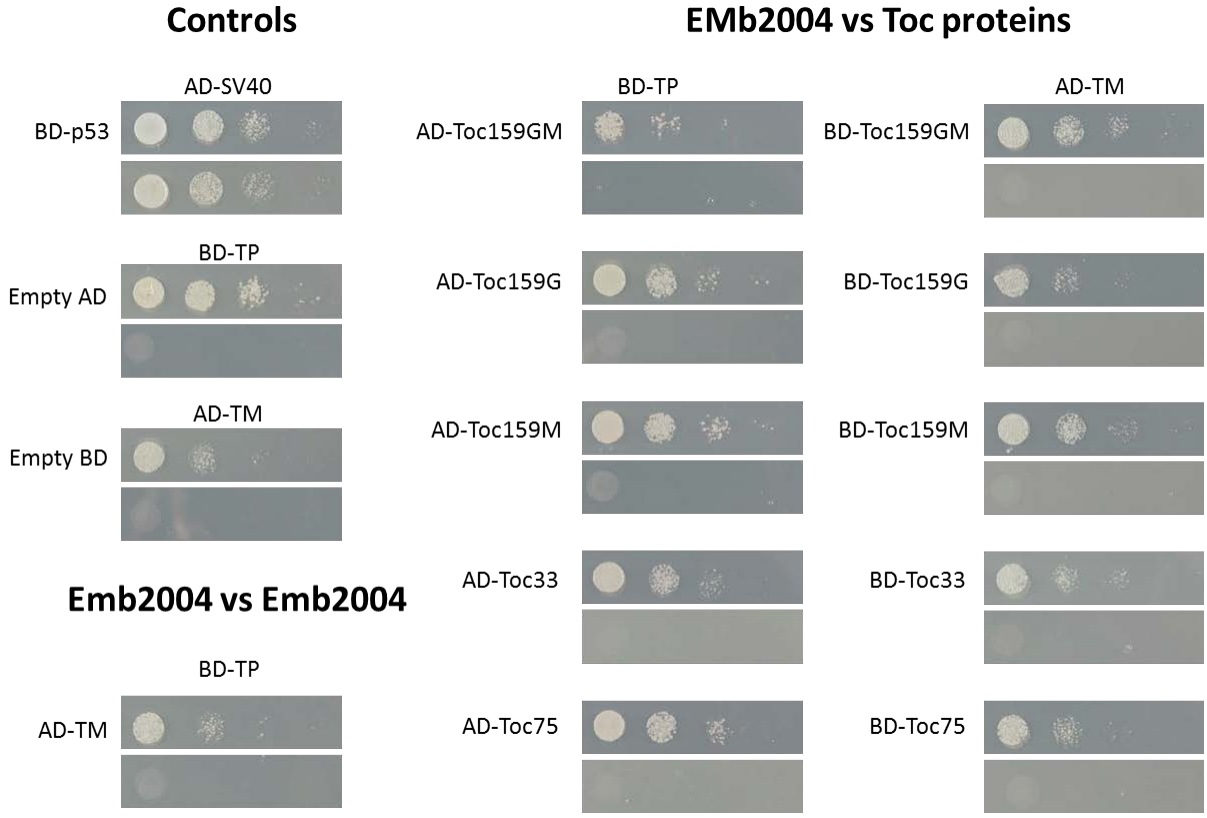


Figure 27: Toc core complex proteins do not interact with Emb2004 in yeast two-hybrid assay: The sequence of Emb2004 without its transit peptide (-TP) was cloned into pGBKT7 to obtain an N-terminal fusion protein with the Gal4 binding domain (BD-TP). The sequence of Emb2004 lacking the transit peptide as well as the transmembrane domain (-TM) was cloned into pGADT7 to obtain an N-terminal fusion protein with the Gal4 activating domain (AD-TM). These two constructs were assayed against each other and against pGBKT7 (BD) or pGADT7 (AD) constructs coding for the Toc core complex proteins, in the AH109 yeast strain. 10^{-1} dilutions of an over-night culture of each of the AH109 lines were spotted on a Trp-Leu drop-out medium to select for the presence of the pGBKT7 and pGADT7 vector respectively (control, upper panels). The same dilutions were spotted on Trp-Leu-Ade-His drop-out medium to select for the activation of the *ADE2* and *HIS3* reporter genes by potential positive interaction between the two fusions proteins tested (lower panels). (Toc159GM: G and M domain of Toc159; Toc159G: G domain of Toc159; Toc159M: M domain of Toc159; Toc33: Toc33 without the 21 C-terminal AA; Toc75: Toc75 without the transit peptide; BD-TP and AD-TM constructs were tested for auto-activation against empty pGADT7 and pGBKT7 vectors respectively. The positive control tested the known interaction between the SV40 large T antigen and the murine p53 protein.

The yeast expression vectors pGBKT7 (fusion with the binding domain of Gal4 and a Nuclear Localization Signal (NLS)) and pGADT7 (fusion with the activation domain of Gal4 and Nuclear Localization Signal (NLS)) were used. The sequence of Emb4004 lacking the transit peptide (Emb2004_TP) or lacking both the transit peptide and the transmembrane domain (Emb2004_TM) were fused to the activating domain of Gal4 or binding domain of Gal4 respectively (resulting in pGADT7_Emb2004_TP (AD-TP) and pGBKT7_Emb2004_TM (BD-TM)). These constructs were co-transformed pairwise with constructs coding for Toc core proteins (Toc33, Toc75 and three constructs with different combination of the Toc159 domains: Toc159GM, Toc159G and Toc159M) fused to the activating or binding domain of Gal4, in the AH109 yeast strain (Clontech). AD-TM and BD-TP were also tested against each other. The AH109 strain is auxotroph for tryptophan, leucine, adenine and histidine and contains three reporter genes, ADE2, HIS3 and LacZ under the control of a GAL4 promoter. 1 to 10^{-5} dilutions of an overnight culture of each of the co-transformed line were spotted on Trp -Leu double drop-out medium (DDO) as a control and on His- Ade-Trp-Leu quadruple drop-out (QDO) medium to test the activation of the reporter genes (fig. 27). The control lines with the AD-TP, AD-TM and BD-TM co-transformed with the empty partner vector grew well on DDO but not on the QDO, controlling for the auto-activation of the Emb2004 construct. The positive control, containing pGBKT7-53 and pGADT7-T (encoding murine53 and the SV40 large antigen respectively) grew well on both media. None of the tested pairwise interaction between Emb2004-TM/Em2004-TP and Toc core complex protein activated the reporter genes. No activation was also shown with Emb2004-TP against Emb2004-TM. However, the absence of interaction in yeast between two proteins does not exclude an interaction in the organism of origin. The fusion of the Gal4 activating or binding domain may mask the interaction site on either one or both the tested proteins. It could also be that the interaction is indirect, the two proteins being part of the same complex. Moreover, yeast is a heterologous system for plant proteins, and it might be that the native physiological conditions or a specific ligand are required for the interaction to occur.

Emb2004-TAP line (Fig. 28):

Emb2004 was not detected in the TEV eluate of theTAP-Toc159 purification which might be due to the incapacity of the anti-Emb2004 antibody to detect the native Emb2004 protein. Therefore the reciprocal experiment was set up. The full length Emb2004 coding sequence was fused at its C-terminus to the sequence of the TAP tag in the pEARLYGATE-TAP plant expression vector, under the control of a CAM35S promoter and with the Octopine Synthase terminator sequence. The TAP tag consists of two ProteinA domain in N-terminal of a Calmodulin Binding domain (CBP) separated by a TEV (Tobacco Etch Virus) protease site. The BAR resistance gene conferring resistance to Glufosinate ammonium allows the selection of positive insertion *in planta* (fig. 28, a). The construct was used to transform *A. tumefaciens* bacteria. Arabidopsis plants (Col2, wt) at the flowering stage were transformed using the floral dip method. Seeds of the dipped plants were grown on soil and sprayed with Glufosinate ammonium. Resistant plants were transferred to individual pots for further characterization. Interestingly 5 plants out the 12 which were resistant to Glufosinate ammonium, starting developing a pale phenotype while reaching the rosette developmental stage (fig. 28, b). To determine the presence of Emb2004-TAP in the 12 transformant lines protein were extracted, separated by SDS-PAGE and subsequently transferred to a nitrocellulose membrane. The nitrocellulose membrane was probed with rabbit IgG to

detect the TAP-tag (fig. 28, c). A band at approx. 90 kDa was detected, higher than the predicted 79 kDa predicted mass of the Emb2004-TAP recombinant protein. However, the 90 kDa band was not detected in the wild type plant extract. Interestingly, the Emb2004-TAP is not detected or to a much lower extent in the extracts of the plants showing a pale phenotype (signaled in orange on the photograph and on the Western blot). The reduction or the complete absence of Emb2004-TAP may be indicative of transgene-induced gene silencing and be responsible for the observed phenotype. The reduction or the complete absence of Emb2004-TAP may be indicative of transgene-induced gene silencing and be responsible for the observed phenotype. The T2 seeds of some of the lines (6, 9, 11, 13) expressing higher levels of Emb2004-TAP were tested for the segregation of the glufosinate ammonium resistance.

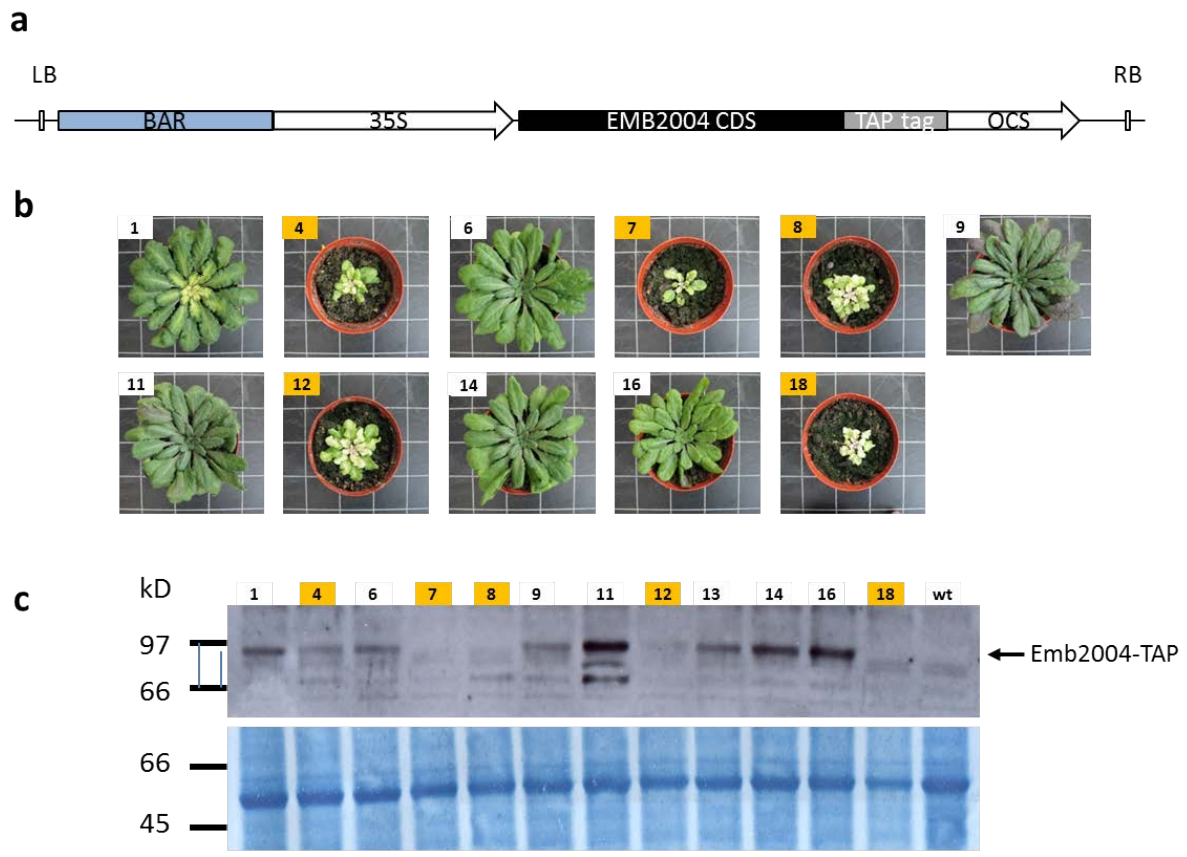


Figure 28 : EMB2004-TAP expressing lines. a) The Emb2004 coding sequence (CDS) was cloned in the plant expression vector pEARLYGATE-TAP with a 35S promoter in 5' and a TAP-tag and Octopine Synthase terminator sequence in 3', to obtain an Emb2004 fusion protein with a C-terminal TAP tag (Emb2004-TAP).The construct was stably transformed in the Col2 wt background using the floral dip method. Positive transformants were selected for the BAR resistance. **b**) Photos of the 11 weeks old resistant plants. Several independent transformant lines displayed a chlorotic phenotype (marked in orange). **c**) 25 µg of protein of the resistant plants shown in **b** were separated on a SDS-PAGE gel and transferred onto a nitrocellulose membrane. The membrane was probed with IgG. An orange box indicates the plants showing a chlorotic phenotype. The arrow indicates the band that corresponds to EMB2004-TAP, which has an expected size of 79.5 kDa without the predicted transit peptide. This band is not present in the control wild type protein extract. The Amido Black staining of the membrane is shown as a loading control.

The distribution of resistant and sensitive plants was compared to the theoretical expected segregation of the resistance for a single insertion event (75% resistant, 25% sensitive) using

a chi-squared test. The lines 9 and 11 followed the expected distribution (9: $p<0.001$; 11: $p<0.05$) suggesting that these line contain only one insertion of the Emb2004-TAP T-DNA construct. Line 11 was kept for further characterization as it showed the highest expression level for Emb2004-TAP and did not develop a pale phenotype. The T3 seeds of different T2 plant individuals of the Line 11 were tested for the resistance to glufosinate ammonium to select a line homozygous for the Emb2004-TAP insertion. 7 out of 20 T2 individuals of Line 11 appeared to be homozygous as all their T3 progenies were resistant to Glufosinate ammonium , consistent with the 1/3 ratio of homozygous plants among T2 resistant plants. Homozygous T3 plants were used to produce seeds.

TAIL analysis of the T-DNA insertion in the Emb2004-Tap plant line (Fig. 29):

TAIL (Thermal Asymmetric Interlaced polymerase chain reaction) was used to localize the Emb2004-TAP T-DNA insertion of Line 11. Six sets of degenerated primers (Each set of degenerated primers allowed for 2 or 4 substitutions at 5 different positions: 3 sets of primers contained 128 different possible sequences, 2 sets 256 different sequences and 1 set, 64) were used in combination with a primer for the left and right border of the T-DNA insertion for a total of 12 different primer pairs. The leaf DNA of Emb2004-TAP Line n° 11 individuals was extracted using the CTAB DNA extraction method. Three successive PCRs were done using the 12 primer pairs described above (The hybridization site of the left and right border primers on the T-DNA was shifted toward the T-DNA border for the 2nd and 3rd round of PCR, providing "nested" PCR products) . A first round of PCR was done on the extracted DNA, the two subsequent rounds on the PCR products of the previous rounds.. The PCR products of the 2nd and 3rd PCR were loaded on an agarose gel and stained with EtBr (fig. 29, a).

An amplification product was present in all 6 of the last PCR reactions on the left border. In some cases, a slightly higher product was already visible in the corresponding 2nd PCR reaction. A PCR product was visible in 5 of the last PCR reactions on the right border, and only one of these products was already visible in the 2nd PCR reaction. The higher and more abundant PCR product of the 3rd round of reaction (1st set of primer for the right border primer and 3rd and 4t^{h} primer set for the left border, indicated with a red arrow on fig. 29) were purified from the agarose gel and sent for sequencing, using the left or right border primer used for the third PCR reactions. The sequences were searched against the NCBI non-redundant DNA sequence collection using the BLAST algorithm. The sequence of the right border product contained only sequence of the T-DNA insertion or *A. tumefaciens* helper plasmid sequence. Both sequences obtained from the left border PCR products contain the sequence of the T-DNA left border downstream of the primer used for sequencing, followed by *A. thaliana* genomic DNA. As the same sequence was obtained with two different degenerated primers, it is likely that it corresponded to a *bona fide* insertion site.

The alignment of the TAIL PCR sequence and the *A. thaliana* genome sequence starts 308 bp upstream of the locus AT5G22520.1 and 960 bp downstream of the AT5G22510.1 locus (fig.29, b), on the plus strand sequence of the chromosome 5. With regard to the right border, it cannot be excluded that it inserted in this AT5G22510.1 locus, but a deletion of more than 900 bp although possible is rather rare. AT5G22510 encodes an alkaline/neutral invertase, an enzyme which converts sucrose to hexose. No visible phenotype is reported for the homozygous mutant of the gene.

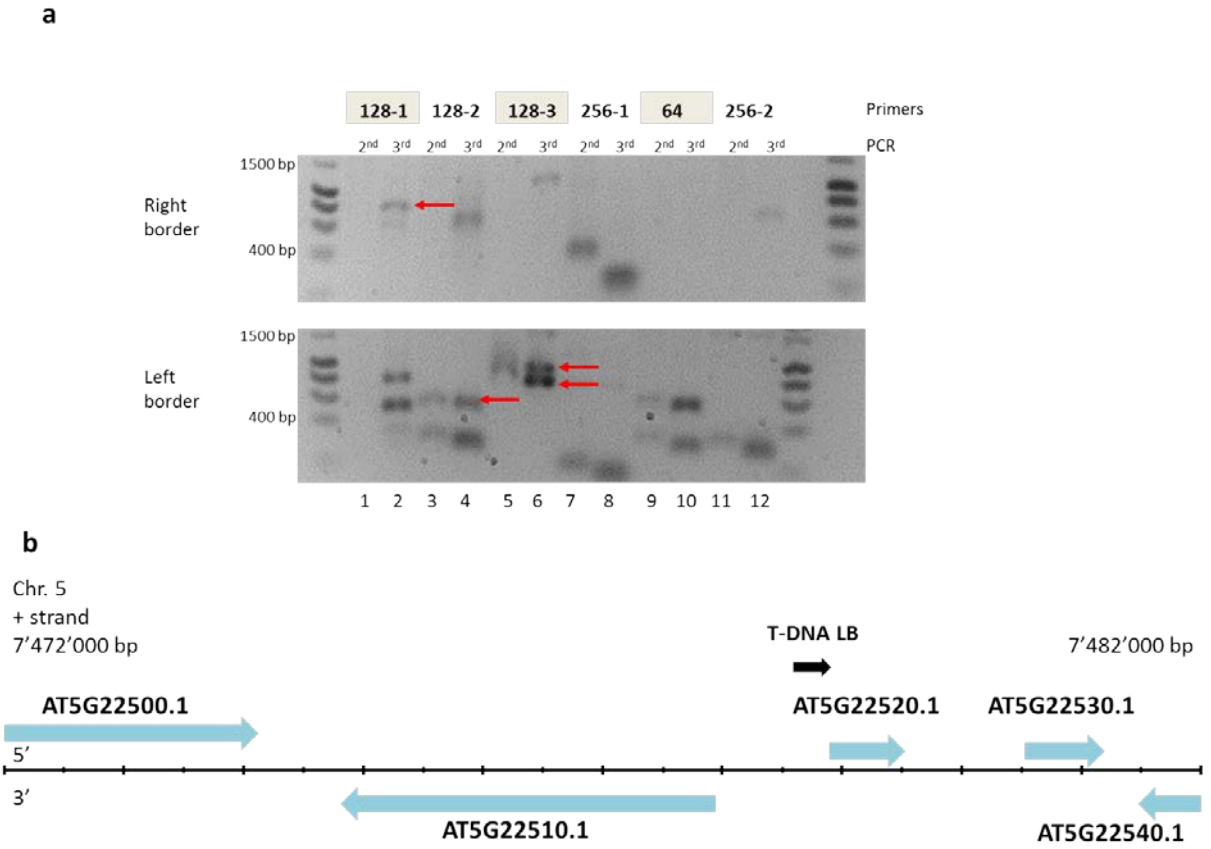


Figure 29: Localization of the T-DNA insertion in the Emb2004-TAP line n°11: a) TAIL PCRs for the right and left border of the T-DNA insertion were done on CTAB DNA of the Emb2004-TAP line n°11. Reactions with 6 different degenerated primers were done for each border (three 128 fold, two 256 fold and one 64 fold degenerated primers were used). The 2nd and 3rd amplification products of each reaction were separated on an agarose gel. The bands marked by a red arrow were excised and sent for sequencing (lane 2, 4, and 6). B) The sequences obtained were searched against the NCBI non redundant nucleotide sequence collection using the BLAST algorithm and mapped on the *A. thaliana* genome. The left border of the T-DNA is inserted in a non-coding region located downstream of the AT5G22510.1 locus and upstream of the AT5G22520.1 locus on chromosome 5 (black arrow: *A. thaliana* sequence identified by TAIL PCR downstream of the T-DNA left border; Blue arrows: protein coding genes). No *A. thaliana* sequence was obtained for the right border of the T-DNA insertion.

Analysis of the silenced line (Fig. 30-31):

Optical and TEM microscopy (fig. 30):

As previously described, different independent lines obtained by the insertion of Emb2004-TAP T-DNA displayed a pale phenotype during development. The lines n° 8 with a pale phenotype (Pale Emb2004-TAP) and n° 18 with a wt phenotype (Emb2004-TAP) were selected for a more careful characterization, wt plant being used as a control (fig. 30, a). Leaves were fixed and embedded in resin for light and transmission electron microscopy (TEM). Toluidine blue stained semi-thin sections and ultrathin sections contrasted with osmium tetroxide and uranyl acetate were used for light microscopy and TEM respectively. Green Emb2004-TAP leaves featured a wild type-like organization with defined palisade and spongy mesophyll as well as epidermal tissues. Chloroplasts were easily distinguishable and

of similar shape in both genotypes. In pale Emb2004-TAP leaves, the cells were larger and less abundant. The spongy and palisade mesophyll were not well defined and chloroplasts were flattened and barely distinguishable from the cell wall (fig. 30, b). The size and shape of the green Emb2004-TAP chloroplasts observed by TEM were similar to the wt chloroplasts. The thylakoid membranes and grana were well visible. In pale Emb2004-TAP cells, no canonical chloroplasts were visible. Their size is approx. half the one of wt chloroplasts and the thylakoid membranes were extremely undeveloped. No grana were visible (fig. 30, c).

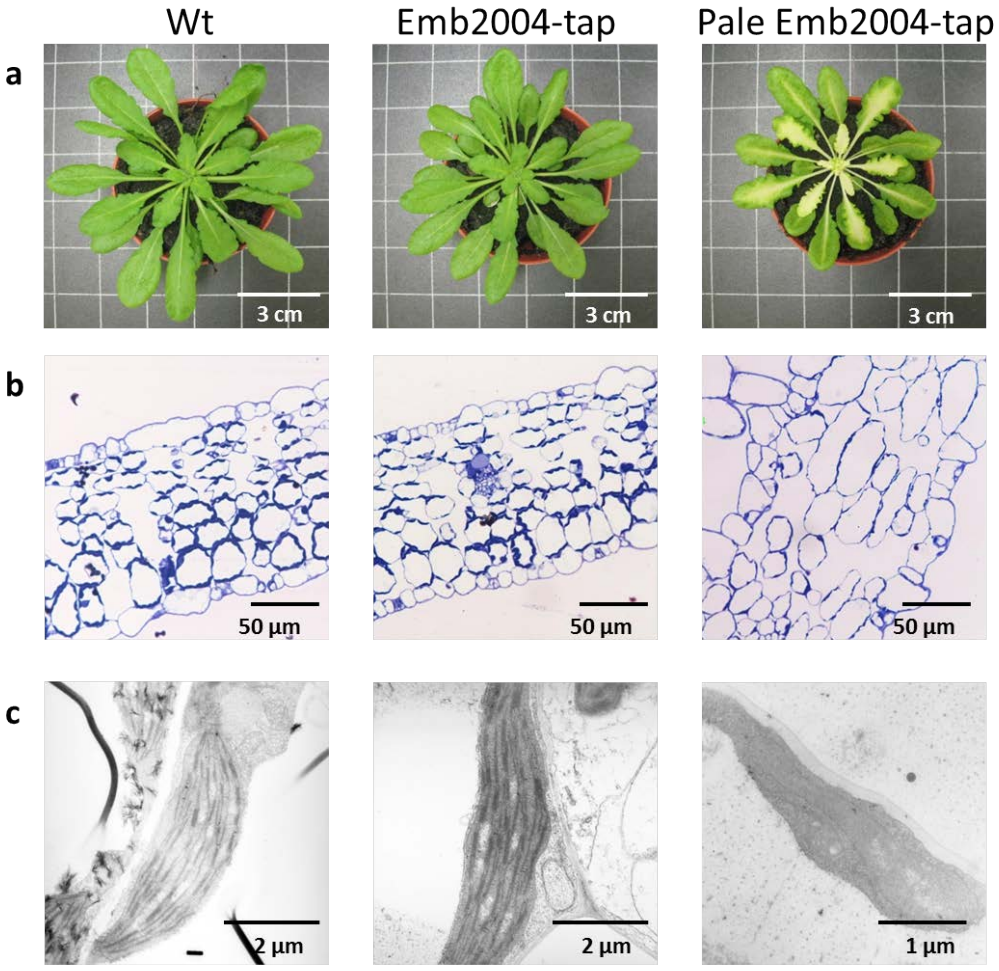
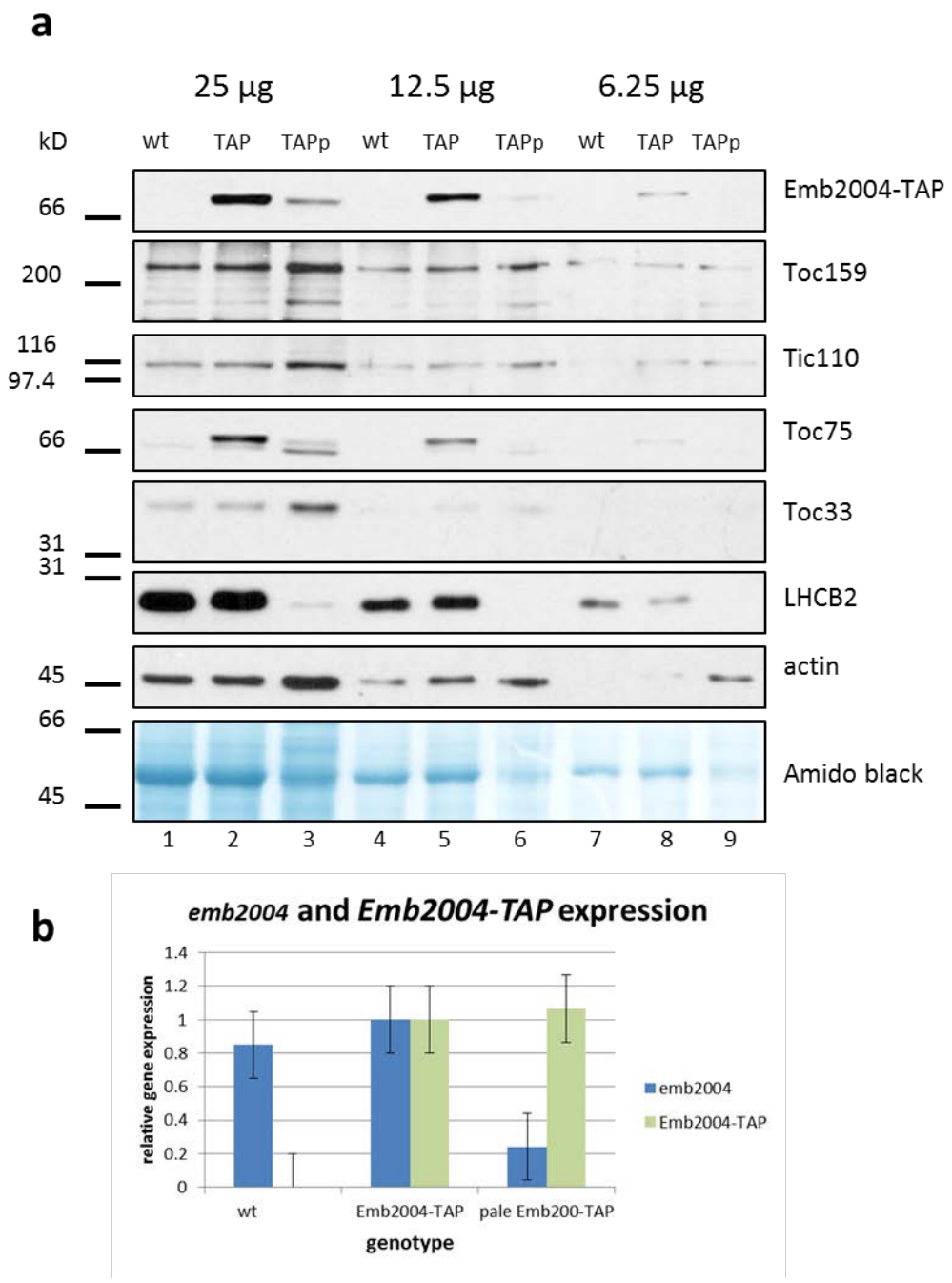


Figure 30: Structure of pale EMB2004-TAP plants : Section of a leaf of a wild type plant, an Em2004-TAP plant and an Emb2004-TAP pale plant **(a)** were observed by light **(b)** and TEM microscopy **(c)**.

Proteins quantification and qRT-PCR quantification of the Emb2004-TAP expression (fig. 31):

Whole leaf protein extracts of rosette stage plant were prepared from the same individuals which were observed by microscopy (fig. 30, a). The protein content was quantified and two different amounts of protein extract (30 and 60 µg) were separated by SDS-PAGE and transferred to a nitrocellulose membrane for immunoblotting (fig. 31, a). Emb2004-TAP was well expressed in the Emb2004-TAP non-affected, green plant compared to pale one. This confirms the previous observations made for the different pale and non-pale T1 plant lines. The same pattern is observed for the abundance of Toc159, Tic110, Toc75, Toc33 and actin.



(Smith et al. 2004)

Figure 31 : Emb2004 pale phenotype may be due to combined transcriptional and translational silencing. a) Equal amounts of leaf protein extracts of the wild type, Emb2004-TAP (TAP) and pale Emb2004-TAP (TAP) shown in figure 27 were separated by SDS-PAGE and blotted onto a nitrocellulose membrane. The membrane was probed with antibodies against the indicated proteins on the left. Emb2004-TAP protein was detected using rabbit IgG. The upper band in the Toc75 panel corresponds to the remaining signal of Emb2004-TAP of a previous probing of the membrane. The Amido-Black staining of the membrane is shown as a loading control. **b)** Quantitative real-time PCR analysis of the expression of the native *Emb2004* and the transgenic *Emb2004-TAP* genes in the wild type, Emb2004-TAP and pale Emb2004-TAP plants shown in figure 27. The primer pairs for the native *Emb2004* and *Emb2004-TAP* genes were designed in the 3' UTR and the TAP-tag sequence respectively. The expression of both genes was normalized with the expression in the Emb2004-TAP plant.

The pale Emb2004-TAP has a slightly higher amount of these proteins than the wild-type and the non-pale Emb2004-TAP. Expression of LHCB2 and RBCS is repressed, both being strongly reduced in pale Emb2004-TAP in comparison to the wild-type and non-pale Emb2004-TAP.These patterns suggest that photosynthetic proteins are strongly down-regulated in pale leaves. The up-regulation of Toc/Tic and the house keeping gene actin is more likely due to the fact that equivalent amounts of protein were loaded. Photosynthetic proteins constitute a huge fraction of the total protein content in green leaves but in pale leaves their level is reduced. Therefore, when equivalent protein amounts of green and pale leaves are loaded, proteins that are unchanged in pale leaves appear overrepresented.

RNA was extracted with a commercial kit from the same plant individuals (fig. 30, a) as for the protein quantification. cDNA were produced and quantified by Real-Time PCR (RT-PCR), using a non-specific DNA probe (SYBRgreen) to measure the expression of the endogenous *Emb2004 gene* and the *Emb2004-TAP* transgene (fig. 31, b). The primers for the endogenous gene were designed in the 3'UTR and the primers for the transgene in the sequence coding for the TAP-Tag. The endogenous *Emb2004* gene expression was equivalent in the wt and the Emb2004-TAP plant, but strongly reduced in the pale Emb2004 plant. Surprisingly, the Emb2004-TAP transgene was equally expressed in Emb2004-TAP and pale Emb2004-TAP plant. A unique RNA extraction and cDNA synthesis was made for each genotype, the RT-PCR was done in triplicate. The expression in the Emb2004-TAP plant was used for normalization.

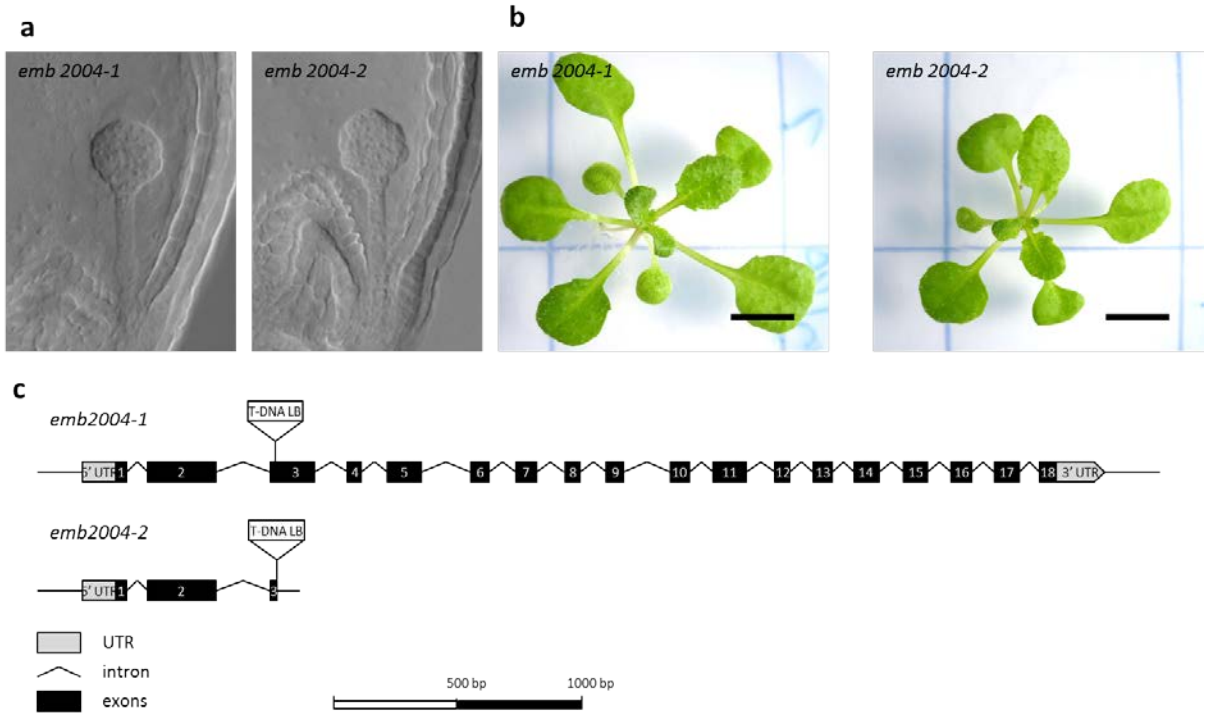


Figure 32 : Emb2004 mutants: *emb2004-1* (Columbia) and *emb2004-2* (Wassilewskija) insertion mutants were isolated by a screen for embryo-lethal mutants by the Meinke lab at Oklahoma State University. **a)** Both homozygote mutants are arrested at the globular stage of embryo development (Photo: SeedGenes project database). **b)** Both heterozygous mutants had no visible phenotype when grown on ½ MS 1% sucrose (scale bar: 0.5 cm). **c)** *emb2004-1* T-DNA insertion is located in the exon n°3, with a 40 bp deletion. *emb2004-2 T-DNA* insertion spans a sequence from exon 3 to the genomic sequence downstream of the 3'UTR, deleting 3500 pp.

Complementation of the *emb2004* mutant (Fig. 32-33):

Two T-DNA insertion mutants of the *Emb2004* locus, *emb2004-1* and *emb-2004-2* were isolated in a genetic screen for embryo-lethal mutations (Tzafrir et al. 2004). Both are arrested at the globular stage of embryo development (fig. 32, a). The heterozygous seedlings have a wt phenotype (fig. 32, b). *emb2004-1* has a T-DNA insertion in the 3rd exon and in *emb2004-2*, the T-DNA insertion deleted a more than 3 kB sequence, from the 3rd exon of Emb2004, to the downstream genomic sequence (fig. 32, c). These mutants were obtained from the ABRC stock (Arabidopsis Biological Resource Center at Ohio State University, USA).

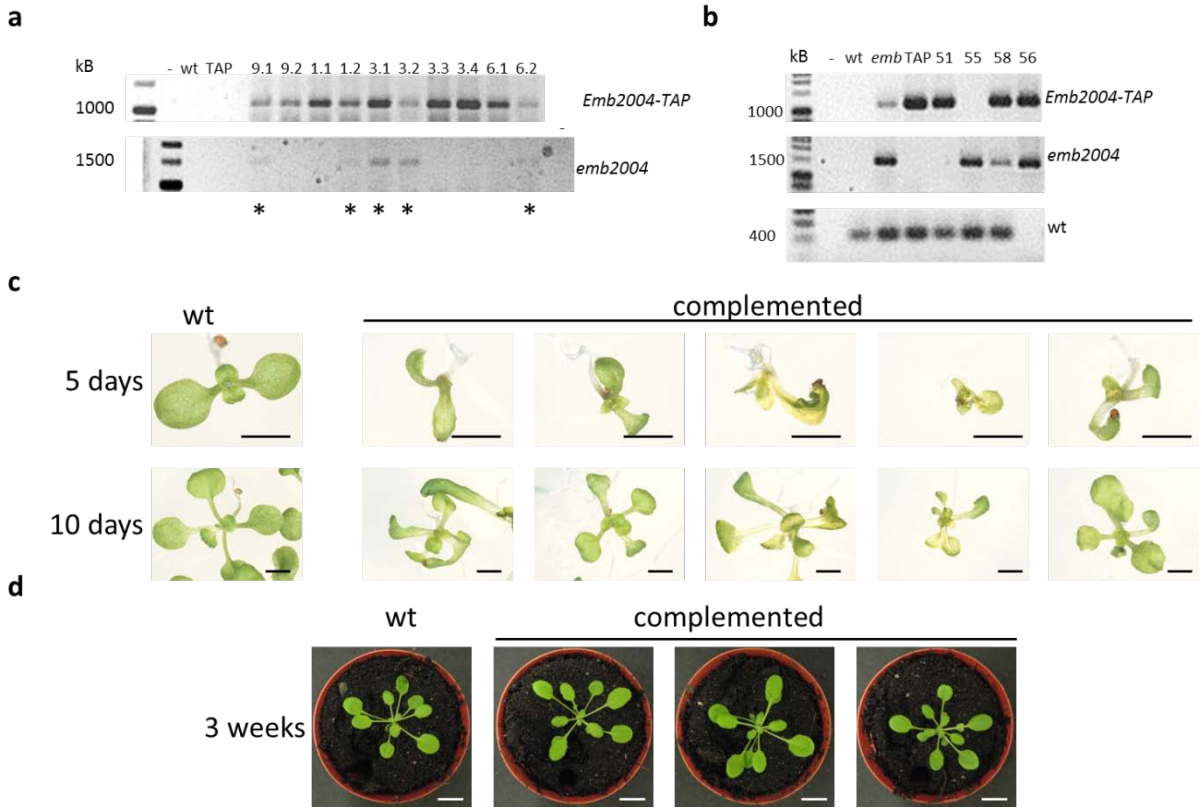


Figure 33: Phenotype of the complemented *emb2004* mutant line: *emb2004* mutants complemented with *Emb2004-TAP* show developmental abnormalities at the seedling stage but have a normal phenotype at the rosette stage. **a)** Agarose gel analysis of the PCR for the genotyping of the first generation of the crosses between heterozygous *emb2004* mutant plants and homozygous *Emb2004-TAP* plants (expressing the Emb2004-TAP recombinant protein in a wt background) (- : negative control; wt: wt DNA; TAP: *EMB2004-TAP* DNA; *Emb2004-TAP*: *Emb2004-TAP* transgene allele; *emb2004*: mutant allele of *Emb2004*). **b)** Agarose gel analysis of the PCR for the genotyping of the progeny of the line 3.2 (heterozygous for the *emb2004* mutation and the *EMB2004-TAP* insertion) shown in **a**. The progeny of the line 3.2 was genotyped for the *EMB2004-TAP* insertion, the *emb2004* mutant allele and the *Emb2004* wild type allele (wt). A complemented line without the wild type emb2004 allele, therefore homozygous for the *emb2004* mutant allele was selected (n° 56). (*emb*: heterozygous *emb2004* mutant DNA). c) Representative seedlings grown on half MS + sucrose of the progeny of the 3.2 line selected in **b** were photographed at 5 and 10 days seedling stage (scale bars: 2mm). **d)** Representative 3 weeks old rosette stage plants of the 3.2 line, in comparison with a wild type plant of the same age (scale bars: 1 cm) (note the individuals are not the same as in **c**).

To assess the functionality of the Emb2004-TAP protein, homozygous *emb2004* mutant plants complemented by the Emb2004-TAP construct were selected. *Emb2004-TAP* homozygous individuals (line n° 11) were crossed with heterozygous *emb2004-1* individuals.

As the *Emb2004* locus is on the chromosome 1 and the *Emb2004-TAP* insertion (line n°11) on the chromosome 5, it is possible to segregate such a complementation. The marker resistance of the *emb2004* mutant and of the *Emb2004-TAP* line is the same (resistance to glufosinate ammonium). The progeny of the crosses was selected on glufosinate ammonium to eliminate non-resistant plants that were wild type for both loci. The resistant progeny was genotyped for the *Emb2004-TAP* allele and the *emb2004* mutant allele.

As expected, all the plants genotyped carried the *Emb2004-TAP* allele and 5 out of 10 individuals carried the *emb2004* mutant allele, consistent with the 50/50 segregation ratio expected (the PCR positive control for the Emb2004-TAP allele failed, but the size of the product obtained in all the other lanes corresponds to the ca. 1000 bp expected size) (fig. 33, a). The progeny of the *emb2004*/*Emb2004-TAP* individuals were genotyped for the Emb2004-TAP allele, the *emb2004* allele and the *Emb2004* allele (wild type allele of *emb2004*) (fig. 33, b). The objective was to isolate an individual homozygous for the *emb2004* allele and homo or heterozygous for the *Emb2004-TAP* allele, corresponding to a ratio of 3/15 resistant individuals when selected for the resistance to glufosinate ammonium. Only one out of 60 individuals genotyped corresponded to the expected genotype. The low yield of complemented individuals might be due to possible false positives in the wt PCR or to a poor complementation rate of the *Emb2004-TAP* construct. Indeed, the progeny seeds of the complemented individual showed a germination rate of only 30%. Moreover, the majority of the 5 day old seedlings of the complementation showed developmental abnormalities (fig. 33, c). Some had three cotyledons, elongated and/or irregular leaf shape and asymmetric growth of the cotyledons. At 10 days, these developmental defects were still visible in the same individuals, but plants at the rosette stage had a normal phenotype (fig. 33, d: please note that these plants are different individuals than the 5 and 10 day old ones). The 35S promoter which controls the expression of the Emb2004-TAP transgene is known to be weakly active during the embryogenesis (Sunilkumar et al. 2002). This might explain the developmental defects observed in cotyledons which develop during embryogenesis, whereas true leaves are formed when the 35S expression of Emb2004-TAP is stronger.

Purification of the Emb2004-TAP protein (Fig. 39):

A first attempt to purify Emb2004-TAP from a whole plant extract, with a protocol similar to the one used for the purification of TAP-Toc159 failed, no Emb2004-TAP being detected in the TEV eluate (not shown). A second strategy, starting with intact isolated chloroplasts gave better results. Intact chloroplast equivalent to 2 mg of chlorophyll were solubilized with TX-100. Insoluble material was removed by centrifugation at 100'000 g. The supernatant was applied multiple times to an affinity chromatography column packed with IgG sepharose beads. A pilot experiment had shown that Emb2004 could not be eluted from the IgG beads using the TEV protease. This is likely due to the conformation of the Emb2004-TAP protein blocking the access to the cleavage site to the protease. Therefore, the IgG beads were

eluted using 0.1 M glycine pH 3. Isolated Col2 chloroplasts were used as the control. The equivalent of 25 µg chlorophyll of the load, an equivalent volume of the pellet fraction of the 100'000 g centrifugation, the flow-through, the first wash, 10% of the last wash and the glycine eluate were separated by SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was probed with IgG to detect the presence of the Emb2004-TAP protein during the different steps of the purification (fig. 34). Emb2004-TAP was depleted from the flow-through fraction, suggesting quantitative binding of the recombinant protein to the IgG column. Emb2004-TAP was well enriched in the glycine eluate. The membrane was probed with antibodies against two Toc proteins (Toc159 and Toc75) and one Tic protein, Tic110. None of these proteins were detected in the glycine eluate of the Emb2004-TAP purification, although they were present in the other fractions.

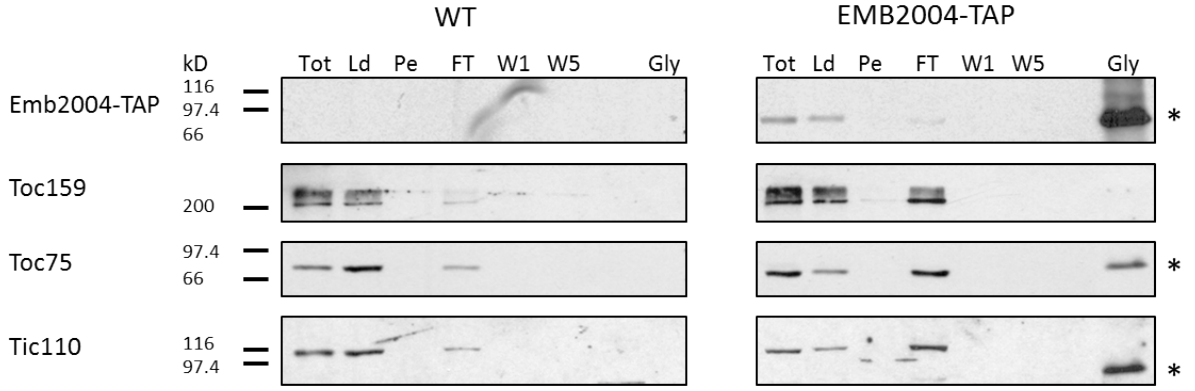


Figure 34: Toc proteins were not co-purified with EMB2004-TAP: Chloroplasts of two weeks old *A. thaliana* plants expressing EMB2004-TAP fusion proteins (line n°11) were isolated. After solubilization in 0.75% TX-100 and centrifugation at 100'000 g, EMB2004-TAP was purified using IgG coated sepharose beads. The IgG beads were eluted with 0.1 Glycine pH 3.0. Samples of the different steps of the purification were separated by SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was probed with IgG to identify Emb2004-TAP and antibodies against Toc159, Toc75 and Tic110 antibodies. Tot: total chloroplast extract; Ld: load of the IgG purification; Pe: pellet fraction after the solubilization with 0.75% TX-100; W1: first wash of the IgG beads; W5: last wash of the IgG beads; Gly: Glycine eluate from the IgG beads. Asterisks indicate the Emb2004-TAP band.

The visible band in the glycine eluate of the Emb2004-Tap purification probed with anti-Tic110 and -Toc75 corresponds to Emb2004-TAP and not to either Tic110 or Toc75. This is because the ProteinA moiety of the TAP tag was not cleaved off and any IgG will therefore detect the Emb2004-TAP protein. A faint amount of Toc159 protein was present in the pellet fraction of the unsolubilized material, but Toc75, another membrane protein was not detected, suggesting that the detergent almost quantitatively solubilized the chloroplast membrane. The membrane was not probed for Toc33, but as the Toc core complex is stable under the detergent condition used, it is unlikely that only this protein would be bound to Emb2004-TAP.

Identification of putative interacting partners of Emb2004:

As no interaction between Emb2004 and Toc159 or other known import complex proteins, two different strategies were followed two identify putative Emb2004 interacting partners. An Emb2004-TAP eluate was analyzed by mass-spectrometry and the BD-TP and BD-TM constructs used to test the interaction of Emb2004 with Toc components in pairwise yeast

two hybrids assays were used to screen a library of *A. thaliana* library in a yeast two hybrids screen (Y2H).

Mass-spectrometry analysis of an Emb2004-TAP eluate (fig. 35):

A glycine eluate as described above was analyzed by mass-spectrometry, using a glycine eluate derived from Col2 plants as a negative control. The glycine eluates were separated by SDS-PAGE and stained by SYPRO ruby to estimate the quality and the quantity of proteins present in the eluate. The gel lanes were cut in three different slices and the proteins therein digested in gel with trypsin. The resulting tryptic peptides were eluted from the gel slices, and analyzed with a LTQ mass spectrometer in triplicates. The spectra of the peptides obtained were searched against "The Arabidopsis Information Resource (TAIR) database 10" containing *A. thaliana* protein as well as common contaminants, using the "Sequest" algorithm. Peptides of contaminants such as human keratin or IgG bleeding from the sepharose beads were detected but other than that the eluates appeared to be of rather good quality (not shown). The protein matches list of the Emb2004-TAP eluate was filtered as follow: PSMs of the three replicates were summed. The proteins which were identified in only one of the triplicates were removed from the list and the protein identified in any of the control triplicates were removed too. The remaining proteins were searched for their curated localization in the At-Chloro database. All were predicted to localize in the chloroplast but proteins predicted only in the thylakoid membranes were removed. The proteins were searched for their "MapManBin" annotation in the Plant Proteome Database (PPDB)(Sun et al. 2009). MapManBins are annotations based on microarrays co-expression studies. It categorize proteins in central metabolic processes and cellular functions (Thimm et al. 2004). By far the most abundant protein (i.e. with the most Peptide Spectrum Matches (PSMs)) was Emb2004, confirming that Emb2004-Tap is correctly expressed in the Emb2004-TAP line n° 11 and was successfully purified (fig. 35). Apart from Emb2004, 4 proteins remained after the filtering. 3 are highly abundant proteins involved in the Calvin cycle or related carbon metabolism (glycerhaldehyde 3-phosphate dehydrogenase A subunit 2 (AT1G12900); Carbonic anhydrase 1 (AT3G01500); ribulose bisphosphate carboxylase/oxygenase (ATCG0049)). The last one is a less abundant "Rieske (2Fe-2S) domain-containing protein" (AT1G71500). The Rieske proteins are component of the cytochrome bc_1 and b_6f complexes at the thylakoid membranes. However the curated At_chloro sub-cellular location indicates a dual location at the thylakoid membranes and at the envelope. This protein might represent an interesting candidate interacting with Emb2004.

Accession	Description	MW (kD)	PSMs	Loc.	MapManBin
AT1G10510	Emb2004	64.72	135	IM	development.unspecified
AT3G01500	carbonic anhydrase 1	29.50	6	Chl	org.transformation.carbonic anhydrases
AT1G12900	glyceraldehyde 3-phosphate dehydrogenase A subunit 2	42.84	5	S/Env	PS.calvin cyle.GAP
ATCG00490	ribulose-bisphosphate carboxylase/oxygenase	52.95	3	Chl	PS.calvin cyle.rubisco large subunit
AT1G71500	Rieske (2Fe-2S) domain-containing protein	31.72	2	Th/Env	misc.other Ferredoxins and Rieske domain

Figure 35: List of the proteins identified by MS/MS in An Emb2004-TAP glycine eluate: Intact chloroplasts were isolated from Emb2004-TAP expressing plants and solubilized with 0.75% TX-100. After centrifugation at 100 000 g, the supernatant was incubated with IgG sepharose beads which were eluted with 0,1 M glycine pH 3.0. Col2 wild type plants were used as a control. The eluates were analyzed in triplicate on a LTQ XL™ Linear Ion Trap Mass Spectrometer mass spectrometer. The proteins identified in two or three of the Emb2004-TAP analysis replicates were kept and cleared from proteins identified in any of the control replicates. The remaining accessions were searched against the At_Chloro database curated for sub-cellular localization and proteins predicted only in the thylakoid were discarded. The PSMs indicated are the sum of the PSM obtained in each replicate. (PSMs: Peptide Spectral Matches; Loc.: curated sub-chloroplastic localization (At_Chloro); MapManBin: functional annotation.)

Yeast two hybrids screen (fig. 36-40):

The pGBKT7-Emb2004 constructs (BD-TP and BD-TM constructs coding for fusion proteins of Emb2004 lacking the transit peptide and Emb2004 lacking the transit peptide and the transmembrane domain with the Gal4 DNA binding domain) used for the pairwise interaction test were used as baits for a yeast two hybrid screen of an *A. thaliana* cDNA library, cloned in the pGADT7 vector. The Clontech matchmaker kit was used. Briefly, the strategy followed included the production of a cDNA library in the AH109 yeast which was mated with an Y187 yeast culture containing the Emb2004 bait protein. Clones with positive interactions were selected on an appropriate drop-out medium. The same procedure was followed in parallel for the two BD-TP and BD-TM bait vectors.

Library construction (Fig. 36-37):

In green tissues, mRNA coding for the photosynthetic proteins like RBCS represent a large fraction of the total mRNA pool. Therefore, RNA from both light grown and etiolated 10 day old seedlings was used for the cDNA library synthesis (fig. 36, a-b). The cDNA was synthetized from RNA using both random primers and oligo-dT primers in separate reactions. Oligo-dT ensures the presence of the 3' part of the sequences and full length sequences in the library. Random primers allow a more homogenous representation of the genes, including internal sequences, no matter what the secondary structure or length of the RNA. Moreover chloroplastic mRNA cannot be reverse transcribed with oligo-dT. The first strand of the cDNA was synthesized by the MMLV reverse transcriptase. MMLV adds overhang of a few guanosine nucleotides in 3' allowing for the hybridization of the primer for the 2nd strand synthesis. Both primers for the 1st and 2nd strand synthesis contain recombination sequence called CDS III and SMART III respectively (fig. 37). The cDNA was purified by gel filtration and co-transformed in the AH109 yeast strain with the pGADT7-Rec2 vector. cDNA recombined *in vivo* with the pGADT7-Rec2 vector through the CDS II and

SMART III flanking sequence. Transformants were selected on Leu single dropout plate, harvested, pooled, concentrated and frozen to constitute the cDNA library. 1/100, 1/1000 and 1/10 000 dilutions of the library were plated on Leu SD to estimate the transformation efficiency. The efficiency obtained was lower than expected according to the Clontech guidelines (10^5 instead of 10^6 clones obtained).

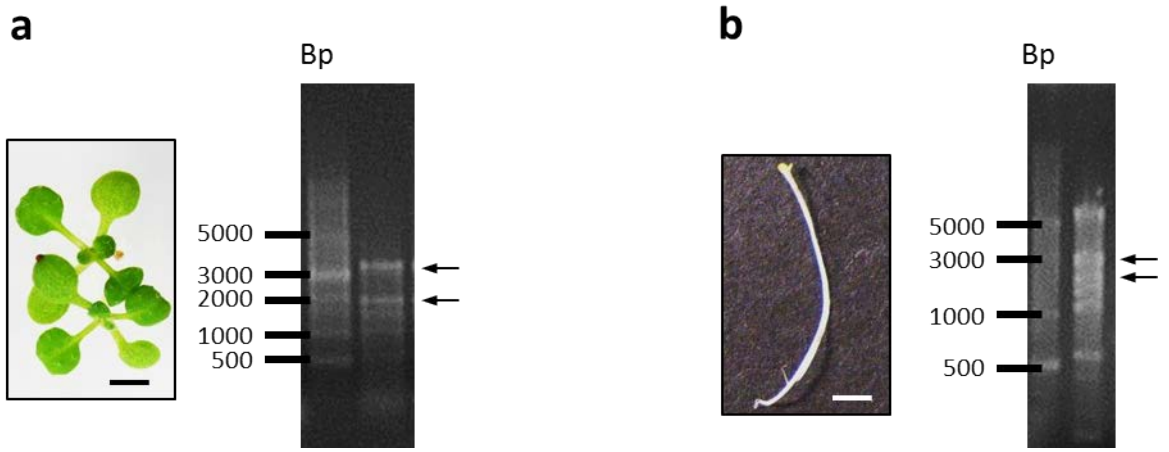


Figure 36 : Construction of the cDNA library for YTH screen: RNAs from 10 old light grown (a) and etiolated seedlings (b) were extracted and loaded on a 2% agarose gel to check integrity (scale bar: 2mm). The 28S and 18S ribosomal RNA bands at 4700 and 1900 bp respectively were well visible demonstrating the integrity of the RNA.

Expression, auto-activation and toxicity tests of the Emb2004 constructs (Fig.38):

BD-TP and BD-TM were transformed in the Y187 and AH109 yeast strain. The constructs were tested for auto-activation (AH109), expression (Ah109 and Y187) and toxicity (Y187). The expression of the recombinant proteins was assessed as follow: Proteins were extracted from yeast cells of an over-night liquid culture and separated by SDS-PAGE electrophoresis. After western blotting, the resulting membrane was probed with anti-bodies against the c-myc tag present in the fusion proteins. The two constructs were well expressed in Y187 in each of the three clones tested. In AH109, only 2 clones for each construct had a detectable expression. The growth curves of the lines containing the BD-TP or BD-TM constructs was similar to the control containing an empty vector, suggesting that the construct have no toxic effects. When streaked on an Ade-His-Leu-Trp QDO plate containing X-α-Gal, AH109 yeast containing either of the Emb2004 constructs or an empty vector were unable to grow. However, a slight blue color was visible around the streaked BD-TP yeasts, suggesting a leaky expression of the LacZ reporter gene. But as the *HIS3* and *ADE2* genes were not activated it was considered that it wouldn't allow false positive colonies to grow during the selection of the positive interaction clones.

Mating of the bait and prey yeast strains:

A fresh culture of AH109 yeast containing the constructs of the bait protein was mixed with an unfrozen aliquot of the Y187 cDNA library and incubated ON at 30 °C for mating. AH109 and Y187 are of the *Mat**a*** and *Matα* mating types, respectively. The diploid yeasts expressing interacting proteins were selected on a Trp-Leu-His TDO plates (fig. 39, a). The mated yeast culture was also plated on Trp and Leu SDO and Trp-Leu DDO to estimate the

mating efficiency which was of 0.054% and 0.161% for the BD-TP and BD-TM respectively, lower than the 2% expected according to the Clontech guidelines. Positive clones were replica plated using filter papers on Trp-Leu-Ade-His X-α-gal QDO for a more stringent selection (fig. 39, b). Although the mating efficiency was low, many positive clones were obtained. Only colonies able to grow bigger and turning blue when transferred to the QDO His X-α-gal were considered positive, i.e. around 150 clones per bait protein. The sequence containing the cDNA of 25 of the colonies of each bait construct was amplified by PCR using the primer provided by the Matchmaker kit manufacturer. 23 and 17 PCR on BD-TP and BD-TM colonies respectively, resulted in a single PCR product (Bands of much lower intensities than the main one were considered unspecific amplification products). 1 and 5 PCRs amplified two products and 1 and 2 PCR amplified no product on BD-TP and BD-TM colony respectively (fig. 40). The amplification of two products is the indication of the presence of two cDNA construct in the yeast clone. The last lane of the gel was loaded with a positive control (pGADT7-SV40 vector). The bands were excised from the gel and purified using a commercial kit and sent for sequencing using one of the primers used for the amplification of the products. The sequences were searched against the NCBI non-redundant nucleotide sequence collection using the BLAST algorithm. The identified sequences are shown in tab. 1 in the annex section. For the BD-TP construct, only 12 sequences were retrieved from the sequencing, and all matched human protein sequence (fibronectin, a protein of the skin). For the BD-TM construct, 20 sequences were retrieved. 4 matched human fibronectin, 9 diverse plant mRNA sequence, and 7 *A. thaliana* sequences, among which 2 rRNA sequence, 2 genomic sequences and 2 mRNA sequences. The sequences matching plants or *A. thaliana* mRNA were searched in the Arabidopsis Information Resource (*TAIR*) database using the BLAST algorithm. However, all the alignments obtained were of only approx. 20 bp, and there were many for one sequence. Even though one of these small sequences could have coded for a true Emb2004 binding domain, it was impossible to know which one of these multiple alignment found for one sequence represented an interesting domain.

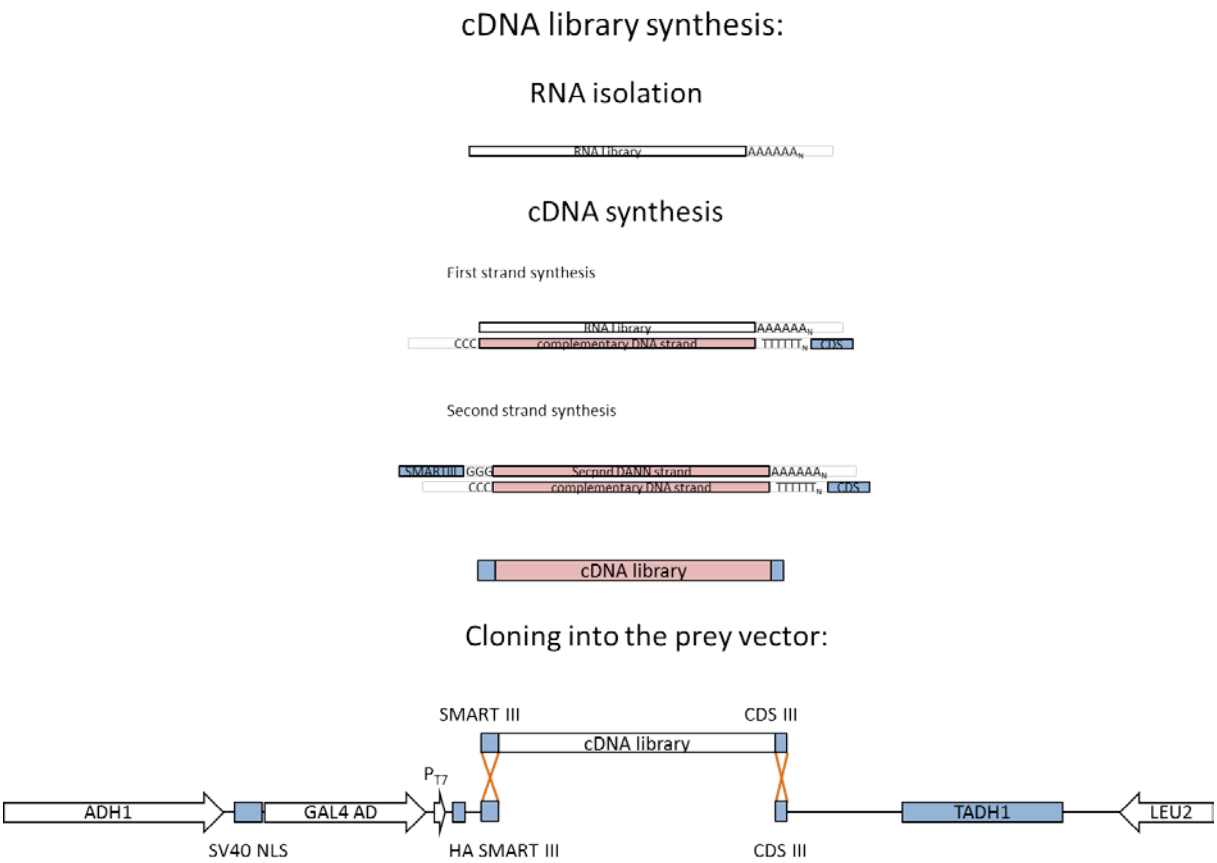


Figure 37 Construction of the cDNA library: The RNAs isolated from light grown and etiolated seedlings were reverse transcribed with the MMLV reverse transcriptase using Oligo dt or random primers flanked with a CDS sequence for priming. The MMLV RT polymerase adds a short cytosine tail in 3' of the first cDNA strand sequence. Second cDNA strand synthesis: A SMARTIII primer with a GGG sequence complementary to the CCC tail synthesized by the MMLV was added and the cDNAs were amplified through 15 cycles of PCR by the TITANIUM Taq polymerase (Clontech). The PCR products were purified by size exclusion chromatography. The cDNAs were co-transformed with the pGADT7-Rec vector in competent AH109 yeast and selected on Trp SDO medium. In AH109 yeast, the SMARTIII and CDSIII sequences recombined with the homologous sequence in the pGADT7 Rec vector.

Ribosomal RNA is one of the most abundant RNA. The use of random primers for 1st cDNA strand synthesis likely explains the presence of these sequences in the cDNA library. The low transformation efficiency obtained when co-transforming the cDNA and the pGADT7-Rec2 plasmid likely resulted in a low library depth, masking the true interaction partners. Ribosomal RNA and collagen related proteins sequences (which is the case of fibronectin) are common contaminants obtained in yeast two hybrid screens (Stephens & Banting 2000). As not one out of 50 clones analyzed gave an *A. thaliana* candidate, it was decided that the remaining positive clones would not be amplified by PCR and sequenced.

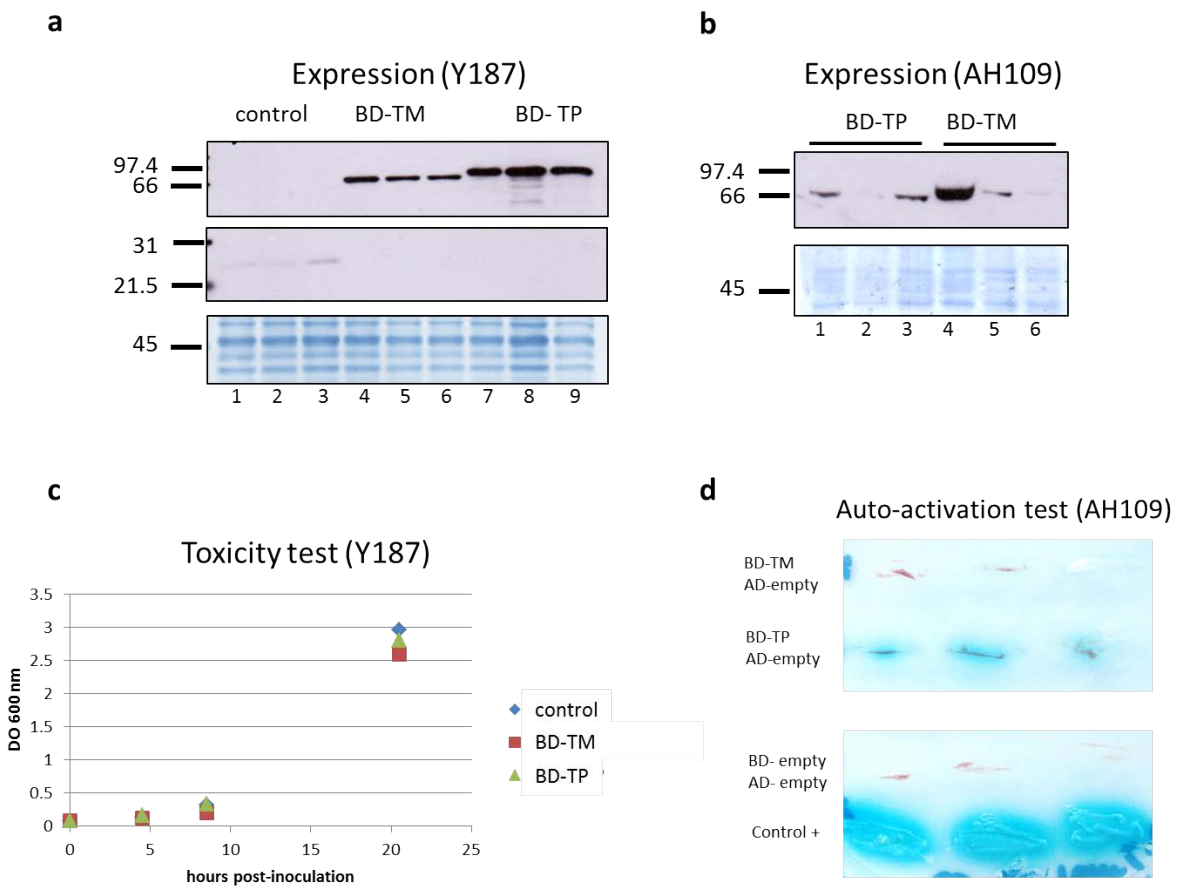


Figure 38 : Emb2004-TP (BD-TP) and Emb2004-TM (BD-TM) constructs are expressed, non-toxic and don't auto-activate: BD-TP and BD-TM construct were transformed in the AH109 and the Y187 yeast strain. **a)** and **b):** Control of the expression of the fusion proteins: Protein of yeast grown in liquid culture were extracted and separated by SDS-PAGE and transferred on a nitrocellulose membrane. The membrane was probed with an anti- c-myc antibody. Both construct were well expressed in Y187 (**a**), only two lines per construct were well expressed in AH109 (**b**). The toxicity of the constructs was assayed in Y187 by monitoring the growth of a liquid culture in drop-out medium. Neither construct impaired the growth of the yeast in comparison to the empty vectors. Yeast growth was monitored by measuring the OD_{600} of the culture. The two Emb2004 constructs were tested for autoactivation in AH109 in presence of the empty activating domain vector. Fresh yeast were streaked on Trp-Leu-Ade-His QDO medium with X-α-Gal. Yeast containing the Emb2004 construct were unable to grow indicating that the His3 and ADE2 reporter genes were not activated. However, a light blue signal was visible around the streaked yeast containing the Emb2004-TP construct, suggesting a leaky expression of the α-Galactosidase reporter gene. The yeast containing the two empty vectors didn't grow and no blue signal was visible. The yeasts containing the positive control construct coding for the SV40 Large T antigen and the murine p53 protein grew well and were able to hydrolyze X-α-Gal.

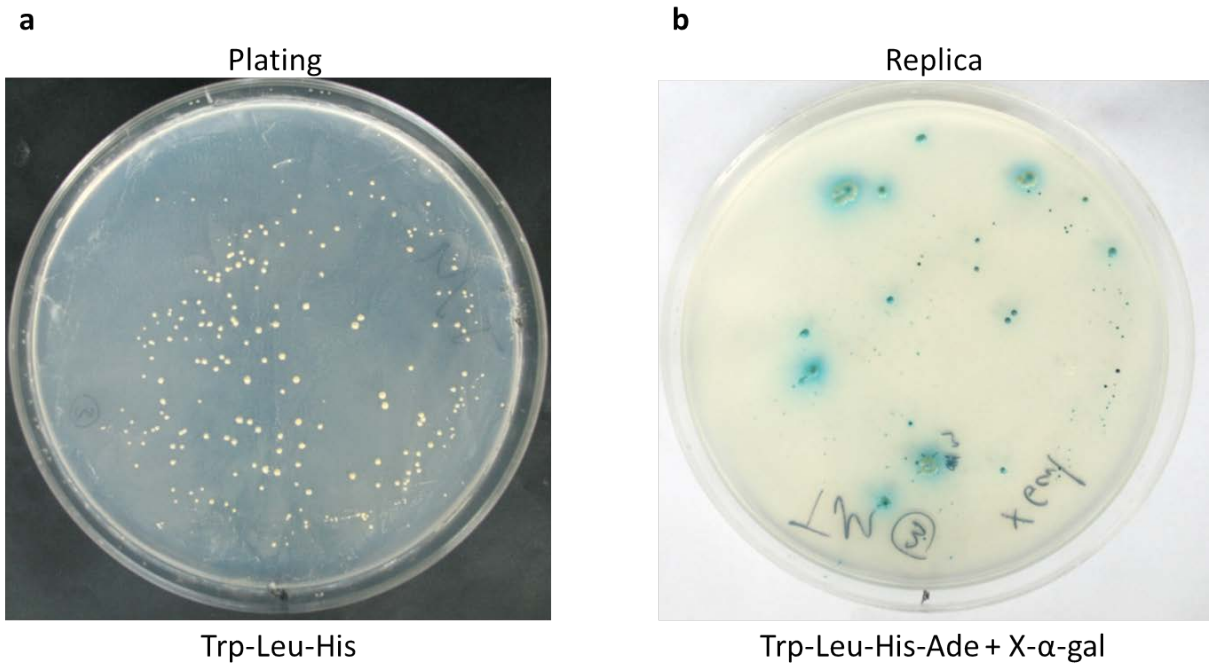


Figure 39: Screen of the cDNA library with the Emb2004-TP and Emb2004-TM construct: AH109 (*Mata*) yeasts containing the cDNA library were mated with Y187 (*Matα*) containing either of the Emb2004 bait constructs. The diploid yeast were selected on Trp-Leu-His TDO selecting for the presence of a bait (Trp) and a prey construct (Leu) and an interaction between the two Gal4 moieties (His). Positive colonies were replicated on Trp-Leu-Ade-His QDO containing X-α-Gal. Only the colonies that were able to grow and displayed a blue color were considered for the sequencing of the prey sequence. The mating efficiency was 0.054% and 0.161%, for Emb2004-Tm and Emb2004-TP respectively, far below the 2% expected.

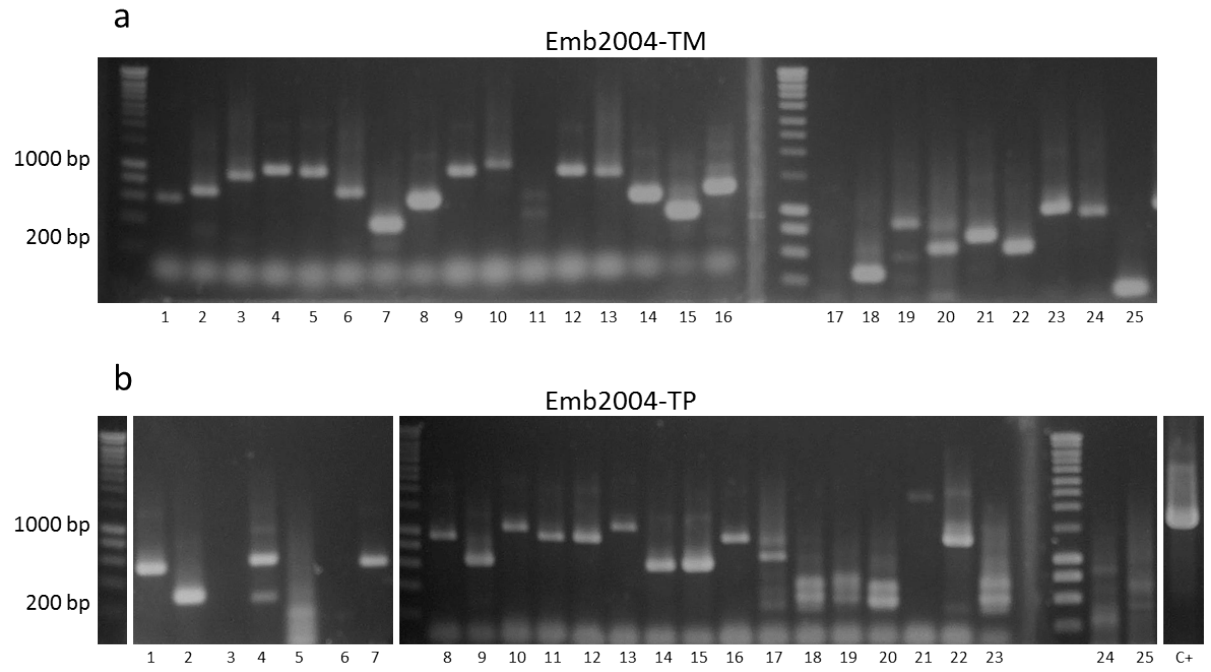


Figure 40: Identification of the sequence of the positive clones: 25 colonies for each screen were analyzed by PCR, to amplify the prey cDNA sequences using two primers flanking the cDNA insertion. The presence of two bands of the same intensities likely indicates the presence of two different prey constructs in the same clone. The bands were excised from the gel, purified and sent for Sanger sequencing, using one of the primers used for the PCR. The pGADT7-SV40 large T antigen vector was used as a positive control.

Outsourced yeast two hybrids screens:

A second screen was outsourced to Hybrigenics, a company specialized in custom yeast two-hybrid screens. For the screen, the Gal4 binding domain was fused to the C-terminus of Emb2004 and screened against an *A. thaliana* cDNA library fused to the LexA activating domain. The screen resulted in the identification of 16 putative Emb2004 interacting proteins (fig. 41). None of these corresponded to one of the proteins identified in the Emb2004-TAP eluate. Each protein was assigned a score based on the number of independent clones identified in the screen (A being the highest score). 3 of the proteins identified in the yeast two-hybrid screen have a chloroplast localization predicted by SUBA, compatible with an interaction with Emb2004, but none have been identified so far in a MS/MS study of the chloroplast (fig. 41). The proteins were assigned a functional annotation (MapManBin retrieved from the Plant Protein Database (PPDB)).

Accession	Description	MW (kDA)	Score	Loc.	MapManBin
AT5G57710	Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein	108.71	A	Chl	stress.abiotic.heat
AT2G26290	root-specific kinase 1	48.09	B	mito	protein.postranslational modification.kinase.receptor like cytoplasmatic kinase VII
AT1G51980	Peptidase family M16	54.40	B	mito	protein.degradation
AT5G64240	metacaspase 3	40.48	B	mito	protein.degradation
AT1G67930	Golgi transport complex protein-related	91.65	B	Chl	cell.vesicle transport
AT5G16210	HEAT repeat-containing protein	132.30	D	cyt	not assigned.unknown
AT3G15160		62.44	D	cyt	not assigned.unknown
AT5G49810	methionine S-methyltransferase	118.71	D	cyt/per	amino acid metabolism.synthesis.aspartate family.methionine.methionine S-methyltransferase
AT1G71040	Cupredoxin superfamily protein	66.16	D	extra	not assigned.no ontology
AT3G01540	DEAD box RNA helicase 1	67.64	D	nuc	RNA.DEAD BOX helicase
AT5G01400	HEAT repeat-containing protein	159.48	D	nuc	not assigned.unknown
AT3G44630	Disease resistance protein (TIR-NBS-LRR class) family	138.98	D	nuc	stress.biotic.PR-proteins
AT3G24715	Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain;	124.86	D	nuc	protein.postranslational modification.kinase
AT1G28320	DegP protease 15	76.12	D	per	protein.degradation
AT5G15710	Galactose oxidase/kelch repeat superfamily protein	50.73	D	PM	development.unspecified
AT5G08720		81.92	F	Chl	not assigned.unknown

Figure 41 : Proteins interacting with Emb2004 identified in an outsourced two hybrid screen: Emb2004 lacking the transit peptide and the transmembrane domain was used to screen an *A. thaliana* cDNA library. Proteins were searched for their predicted location in SUBA (loc.) and assigned a functional annotation (MapManBin). The score is based on the number of different clone identified in the screen for a proteins, A being the highest score.

Discussion:

This thesis focuses on the identification and initial characterization of new components of the chloroplast protein import machinery. To achieve this objective, we affinity-purified N-terminally TAP-tagged TOC159 (TAP-TOC159) from isolated and detergent solubilized chloroplasts. Isolated TAP-TOC159 together with potential interacting factors was initially identified by mass spectrometry. This thesis presents three such proteins, AT4G32250/ AT4G32250, Tic56/AT5G01590 and EMB2004/AT1G10510.

AT4G32250 kinase:

AT4G32250 is a predicted protein kinase with a predicted C-terminal transmembrane helix. Although the SUBA database predicts AT4G32250 to be cytosolic, 3 independent proteomic studies identified it in the chloroplast envelope (Ferro et al. 2010). We therefore hypothesized that AT4G32250 may be one of the kinases phosphorylating the A-domain of Toc159. To do so, AT4G32250 would be expected to be at the outer membrane of the chloroplast facing the cytosol. For *in vivo* localization, YFP was fused to the C-terminal of AT4G32250. This fusion protein resulted in fluorescent labeling at the chloroplast periphery (fig. 12) consistent with chloroplast envelope localization. By confocal microscopy, however, the precise localization in the outer or inner envelope membrane cannot be determined. Thus additional experimentation such as a protease sensitivity assay is required. In addition, the predicted transmembrane helix of AT4G32250AT4G32250, like the fused YFP, is at C-terminus of the protein. Therefore, it cannot be excluded that YFP may have somehow perturbed targeting or insertion of AT4G32250. In *in vitro* import/insertion experiments, AT4G32250 was sensitive to thermolysin after membrane insertion indicating that AT4G32250 is exposed at the cytosolic surface of the outer envelope (fig. 11). The integral or peripheral nature of the interaction of AT4G32250 with the envelope remains to be investigated, either by testing its resistance to alkaline $NaCO_3$ extraction, or by a chloroplast insertion experiment with a radiolabeled AT4G32250 lacking the predicted C-terminal transmembrane domain. However, the prediction of a C-terminal hydrophobic transmembrane helix strongly suggests that AT4G32250 is an integral protein.

The presence of AT4G32250 at the outer envelope and its interaction with Toc159 makes it a good candidate as a kinase for the Toc159 A-domain (fig. 14). However, the NCBI BLAST algorithm predict a conserved Protein kinase B (PKB) domain in the AT4G32250 sequence and no PKB phosphorylation site are predicted in the Toc159 sequence (NetPhosK) (Blom et al. 2004). The prediction also suggests that AT4G32250 don't correspond to the CK2 like activity highlighted in the chloroplast envelope (Agne et al. 2010). The absence of a predicted PKB phosphorylation site does not exclude the presence of a phosphorylation site for AT4G32250. Indeed, the NetPhosK prediction method used is based on experimentally verified phosphorylation sites of a small number of model kinases suggesting that the prediction might not represent the full phosphorylation spectra of a kinase family (Blom et al. 2004). *In vitro* or *in vivo* experiments would be needed to address the kinase activity of AT4G32250 on the A-domain. In vitro phosphorylation assays with recombinant A-domain and AT4G32250 are currently being carried out by Monica Arias in the Kessler Lab. Alternatively, the kinase activity of the chloroplast envelope fraction of wild type on

recombinant A-domain might be assayed using the AT4G32250 homozygous mutant as negative control. Phosphoproteomics on a triple mutant of *A. thaliana* SnrK2 kinase resulted (P Wang et al. 2013) in the identification of substrates including Toc159. Phosphoproteomics would broaden the spectra of AT4G32250 potential targets beyond Toc159. Even though the AT4G32250 null mutant visually has a wild type phenotype it cannot excluded that it has reduced import activity. This has proven to be the case for different null mutants or complemented plants carrying mutant versions of Toc/Tic components.

SnRK2 kinase phosphorylates Toc159 at T692 which is in the A-domain. Moreover, Toc120 and Toc132 are also targets of SnRK2. The snrk2.2/2.3/2.6 triple mutant is insensitive to ABA, suggesting that ABA signaling influences protein import. ABA signaling is implicated in seed dormancy and response to environmental stress such as pathogen attack or drought. However, the effects of the phosphorylation of the Toc159 A-domain and the import process by AT4G32250 or other kinases such as cytosolic CKII currently remain obscure. As mentioned in the introduction, the A-domain of Toc159 and its homologues are likely involved in the selectivity of the import apparatus towards client proteins. Phosphorylation might affect the selectivity of the Toc translocon by altering the biochemical properties of the A-domain by adding negatively charged phosphate groups. The A-domain is already acidic, i.e. it is constitutively highly negatively charged which would be further increased by phosphorylation. It can be speculated that the charge of the A-domain might play a role in selectivity toward pre-proteins through the interaction with transit peptides that tend to be positively charged. It is imaginable that phosphorylation at the A-domain may further modulate A-domain interaction. Alternatively, phosphorylation of the A-domain may play a role in regulating the stability of Toc159, for instance by targeting Toc159 for degradation by the ubiquitin proteasome system (UPS). Toc159 and other Toc proteins have been shown to be targets of UPS, mediated by the SP1 E3 ligase. Phosphorylation is known to promote targeted ubiquitination and degradation by the 26S proteasome (Trujillo & Shirasu 2010).

AT5G01590/Tic56:

This study reports the first evidence for a direct interaction between proteins of the Toc core complex and Tic56. This is highly relevant because it suggests the existence of large complex that extends even beyond the 1 MD mass of the complex identified by Kikuchi et al. (2013). By Western blotting we have only confirmed the presence of Tic56 and Tic20 in the TAP-Toc159 complex. However, by mass spectrometry we also detected Tic214 and Tic100 (tab. 1) which still needs to be confirmed by Western blotting with appropriate controls.

Kikuchi and colleagues identified Toc159, Toc33 and Toc75 but not Tic110 and Tic40 associated with a purified pSSU import intermediate, together with the members of the 1 MD Tic complex (Tic214, Tic100, Tic56 and Tic20) (Shingo Kikuchi et al. 2013). Moreover, the BN-PAGE analysis of the 1 MD Tic complex by the same group showed that neither Tic110 nor Tic40 are present in this complex. However, Kessler et al. (1994) discovered the Toc core complex proteins as well as Tic110 and an unidentified 36 kDa protein (maybeTic40) associated with the late pSSU-ProtA import intermediate, but none of the proteins of the 1 MD Tic complex were present as visible bands. Kouranov et al. (1998) highlighted the existence of a Toc/Tic supercomplex containing the three Toc core complex proteins, Tic110,

Tic22, Tic and the stromal ClpC and cpn60 in the absence of a pSSU import intermediate and Nielsen et al. (1997) co-purified Tic110 with Toc75 in the absence of a precursor protein.

The absence of bound Tic110 and Tic40 with a pSSU import intermediate in the Kikuchi et al. (2013) study could be explained by the use of different detergents (digitonin was used to solubilize the chloroplast envelopes whereas in the present study and the others cited above Triton X-100 was used). At least two hypotheses can account for the association of Tic110 and Tic40 and the proteins of the 1 MD Tic complexes with TAP-Toc159 highlighted in this study. 1) The Toc core complex may associate with both Tic110 and Tic40 and the 1 MD Tic complex (or at least Tic20 and Tic56) to form one Toc/Tic supercomplex (fig. 18, b) or 2) two populations of Toc/Tic supercomplexes coexist, one constituted of a sub-fraction of the Toc core complex together with Tic40 and Tic110, the other of a sub-fraction of the Toc core complex with the 1 MD Tic complex (or at least Tic20 and Tic56) (fig. 18, **a**). This would extend to at least Tic56 the conclusion of (Kouranov et al. 1998) who proposed that Tic110 and Tic 20 indirectly interact through their common association with the Toc core complex. Naturally, as Tic110, Tic40, Tic56 and Tic20 are far less enriched than Toc33 and Toc75 in the TAP-Toc159 eluate, only a sub-fraction of the Toc core complex can be isolated as part the Toc/Tic supercomplex. Kikuchi et al. (2013) questioned the role of Tic110 and Tic40 as protein import components, and based on the phenotype of Tic20, Tic56, Tic100 and Tic214 mutants proposed the 1 MD Tic complex as a general complex for protein import at the inner membrane. However, Köhler et al. (2015) showed by proteomics that many different proteins are still imported and processed in the Tic56-1 null mutant. The pale Tic56-3 mutant expressing a truncated form of Tic56 was shown to have import efficiency similar to a wild-type plant. It is also well known that the complete block of import arising from the knock-out of Toc75, a unique and universally accepted component of the import machinery, leads to embryo lethality. These elements question the role of the 1MD complex as a general translocon and further support the existence of alternative routes. The fact that phenotypes of *ppi2* (lacking Toc159) and *tic56* are both albino but not embryo lethal, suggests that they may cooperate in a pathway that is heavily (but not exclusively) implicated in the import of photosynthetic proteins.

Emb2004:

Localization:

The putative Toc159 interaction partner this thesis deals with is EMB2004. EMB2004 was identified with a high number of peptide by mass spectrometry of TAP-Toc159 complexes. EMB2004 is a leucine rich repeat (LRR) protein with homology to RAN GTPase activating protein (RAN GAP). Leucine Rich Repeat motifs are present in protein of various unrelated function and their common characteristic seems to be the ability to offer a binding site for protein-protein interaction (Bernoux et al. 2011). GTPase activating proteins are essential to complete the GTPase cycle of many GTP-binding proteins. We therefore hypothesized that EMB2004 is a candidate Toc159 GAP. The exciting characteristics and hypotheses made Emb2004 an interesting choice for further characterization.

An antibody raised against Emb2004 was unable to unambiguously identify Emb2004 in a plant extract. Therefore, the localization of Emb2004 depended on other evidence: the Emb2004-YFP localization, the in vitro import experiment, the protease protection assay

together with the alkaline Na_2CO_3 extraction assay provided convergent evidence supporting the localization at the chloroplast inner envelope membrane. A construct encoding Emb2004-TAP complemented the embryo lethal phenotype of *emb2004* null mutant indicating that the recombinant fusion protein was functional. In two parallel experiments, both the Emb2004-TAP protein and imported radiolabeled Emb2004 showed identical association with the membrane fraction of the chloroplast and the same resistance to the alkaline Na_2CO_3extraction. The non-processed form of Emb2004 visible in the import experiments was resistant to alkaline Na_2CO_3. Alkaline extraction detaches protein which are not linked to a membrane by hydrophobic interactions, but an early import pSSU was shown to be resistant to alkaline extraction (Waegemann & Soil 1991). The conclusion is that the *in vitro* imported radiolabeled mature Emb2004 is correctly localized at the chloroplast envelope validating its use for the protease protection assays to determine sub-organellar localization. Moreover, the curated AT_CHLORO subcellular localization of Emb2004 based on MS/MS analysis is also at the inner envelope membrane of the envelope (Ferro, 2010; Bruley, 2012)(Ferro et al. 2010). The YFP localization indicates that Emb2004 is likely not targeted to the mitochondria. NbCEP1, the homologue of Emb2004 in Tobacco was also shown to be localized at the chloroplast envelope using GFP localization (Jeon et al. 2010).

Failure of demonstrating that Emb2004 to interact with Toc159:

We attempted to confirm the interaction of Emb2004 with the Toc core complex using independent and complementary methods. The absence of interaction between Toc159 and Emb2004 observed by yeast two-hybrid assay and immunoblotting on TAP-Toc159 and Emb2004-TAP eluates contradict the preliminary mass-spectrometry identification of Emb2004 in the TAP-Toc159 eluate.

The failure to detect Emb2004 in the TAP-Toc159 complex by Western blotting may reflect the very low concentrations of Emb2004 in the eluate. In this case the sensitivity of the antibody may well be insufficient to detect the small quantities of Emb2004. As already mentioned, it was not possible to unambiguously identify Emb2004 in plant extracts and it cannot be excluded that the quality of this antibody was simply not sufficient.

The spectral counting method which was used for the mass spectrometric isolation is considered semi-quantitative. The abundance of a given protein in a sample correlates quite well with the number of spectra obtained. However, at equimolar concentrations, two different proteins will yield a distinct number of spectra, depending on the number of potential tryptic peptides. The physico-chemical characteristics of a peptide will also influence the detection by the MS/MS. The APEX method used here accounts for the number of predicted tryptic peptides of a protein. The APEX value of each protein identified in the eluate was divided by APEX value of the same protein in whole leaf protein extract thereby controlling for the potential over- or underrepresentation of Emb2004 by MS/MS. However, the analysis of the Emb2004-TAP eluate was negative for Toc159 and other Toc/Tic components by Western blot as well as MS/MS analysis.

Regarding the failure to detect interaction between Toc159 and Emb2004 in the Yeast-Two hybrid assay, the conditions may have been too stringent to detect an interaction between Toc proteins and Emb2004 or the tertiary structure of the Emb2004-TP/TM fusion with Gal4

BD or AD may have prevented a potential interaction. However, the localization of Emb2004 at the stromal face of the inner envelope membrane is not readily compatible with a direct interaction of Emb2004 with the members of the Toc complex and with a GAP (GTPase Activating Protein) activity on Toc159 or Toc33 GTPase domain. However, it can't be completely excluded, based on the mass-spectrometry results that Emb2004 interact indirectly with Toc159 via other components of the import machinery at the inner membrane of the envelope.

Complementation of the emb2004 mutant by Emb2004-TAP

To characterize Emb2004 and identify interaction partners, the *emb2004* mutant was complemented by a T-DNA construct encoding Emb2004 carrying the Tandem Affinity Purification (TAP) tag at its C-terminus.

We attempted to determine the genomic localization of Emb2004-TAP insertion by TAIL PCR. The Ti plasmid sequence obtained by the sequencing of the product of the right border TAIL PCR was somewhat surprising. It is not excluded that a recombination event occurred in the *A. tumefaciens* bacteria between the T-DNA plasmid and the helper plasmid which was later integrated in the *A. thaliana* genome. It could also be that this sequence was amplified from a contamination coming from *A. tumefaciens* plasmid DNA during the DNA extraction or the TAIL PCR. It is known that truncation of the right border of the T-DNA during integration into the *A. thaliana* genome often occurs. This might explain the overall lower number of PCR products in the right border PCR reactions and the fact that only one product was detected in the 2nd PCR reactions. However, it is likely that the *Emb2004-TAP* T-DNA insertion did not disrupt a gene, as the position of the insertion of the left border was located in genomic sequence quite distant from the next coding sequence. Moreover, no phenotype was observed. The complementation of *emb2004* mutant shows that the Emb2004-TAP protein used for the sub-chloroplastic localization and mass-spectrometry for the identification of interacting partners is functional.

Pale phenotypes and undeveloped chloroplast observed in Emb2004-TAP lines:

A pale phenotype was observed in 5 independent lines. It is therefore unlikely that in each case the T-DNA insertion disrupted a gene resulting in the phenotype. The pale phenotype observed in these Emb2004-TAP lines is most likely due to transcriptional silencing also known as co-suppression. Emb2004-TAP was clearly less abundant in pale plants. Endogenous Emb2004 mRNA was down-regulated but not that of the Emb2004-TAP transgene. Emb2004-TAP RNAs might be less sensitive to targeted degradation or its expression under the 35S promoter might ensure the replacement of the degraded RNA. The unchanged expression of Emb2004-TAP RNA together with down-regulation of the Emb2004-TAP protein abundance suggests that translational silencing may have occurred (Baulcombe, 2004).

Expression analysis using the GENEVESTIGATOR resource shows that Emb2004 is expressed more or less constantly during the development of the plant. The pale phenotype due to silencing is present at the adult stage. This suggests that Emb2004 is not only required for plastid function during embryogenesis. However, the embryo lethal phenotype, the poorly developed chloroplasts in pale silenced lines and the fact that the albino phenotype appears

at the center of the rosette support a role for Emb2004 at the early stages of chloroplasts biogenesis. But it cannot be excluded that silencing is present only in the center of the rosette. Similar results were obtained for the homologue of Em2004 in *N. benthamiana* NbCEP1 where virus induced silencing induced a pale phenotype and underdeveloped chloroplasts (Jeon et al. 2010).

Purification of Emb2004-TAP and MS analysis:

Emb2004-TAP was purified from transgenic plants using IgG affinity chromatography. However, the yield of peptides identified other than Emb2004 in the Emb2004-TAP glycine eluate was very low. Emb2004 had 135 peptide identified whereas carbonic anhydrase 1 (AT3G01500), the next protein on the list had only 6 peptide detected. However, it can be argued that there was too much prey for the potential baits as over-expressed Emb2004-TAP might have been more abundant than the endogenous protein. It might be that the interaction of Emb2004 with a partner depends on the presence of a ligand or specific physiological condition. If only a weak or transient interaction occurs, chemical cross-linking would be required to capture potential interacting partners. The proteins identified in the Emb2004 eluate might well be true interactors, but extreme caution is recommended. 3 out of the 4 proteins identified are highly abundant proteins implicated in Calvin cycle or general carbon metabolism and are likely contaminants (fig. 35). The last hit is a "Rieske domain containing protein", a less abundant protein predicted to localize at both the envelope and the thylakoid. However, Rieske proteins are subunits of the mitochondrial cytochrome bc1 complex or the b_6f complexes at the thylakoid membranes. The presence of this protein at the envelope might be linked with its folding and not with its function. Indeed, Madueno et al. (1993) showed that an *in vitro* imported Rieske Iron-Sulfur protein forms a complex with both the Cpn60 complex and Hsp70. Only two peptides were identified in the three analytical replicates, which is extremely low, suggesting that this "Rieske domain containing protein" is likely a contaminant.

Yeast two hybrids screens:

As an additional technique to identify potential Emb2004 interacting proteins an in-house yeast two-hybrid screen was carried out. The absence of interesting *A. thaliana* genes in the sequences identified in the yeast positive clones of the Emb2004-TP and Emb2004-TM screen is likely due to the poor depth of the cDNA library rather than the absence of potential Emb2004 interacting partners. It is also possible that the tertiary conformation of the fusions of Emb2004-TP/TM with the Gal4 binding obstructed a potential protein binding domain. In hindsight, it may have been a good idea to screen the library with Emb2004 fused to the Gal4 binding domain in N-terminal, in addition to the C-terminal fusions that were used.

The outsourced yeast two hybrid screen using Emb2004 lacking the transit peptide and the transmembrane domain was more fruitful. However, only 3 out of the 16 proteins identified have a predicted location at the chloroplast.

AT1G67930 is a "Golgi transport complex protein-related" involved in Golgi retrograde vesicle trafficking in HeLa Human cell (Pokrovskaya et al. 2011). Vesicular trafficking in chloroplast remains largely unknown but two GTPases were shown to be involved in this

process (Karim & Aronsson 2014), suggesting that AT1G67930 might be an interesting potential Emb2004 interacting partner to investigate.

AT5G08720 has a Cyclase/dehydrase domain (Interpro database, (Mitchell et al. 2014)) involved in the synthesis of polyketide secondary metabolites in *Streptomyces spp*. But this domain, although conserved in different organisms remains largely unknown. Moreover, it has the lowest score of the list of the proteins identified in the screen (F) suggesting that it would represent a risky candidate to investigate.

AT5G57710 or SMAX1 is a suppressor of mutants of the MAX2 pathway which is involved in the control of plant growth (Waters et al. 2014). Interestingly, it belongs to a family of 8 genes which are related to Hsp101, a ClpB chaperone involved in heat tolerance (Young et al. 2005) and interacting with Hsp70 in Maize (Zhang & Guy 2005). Chaperones are thought to be involved in protein folding at the inner envelope of the chloroplast (see introduction), suggesting that this protein might represent an interesting candidate to investigate regarding the localization of Emb2004. The potential interaction of these 3 proteins with Toc159 or Toc core complex component would need to be investigated in *A. thaliana* as interactions in Yeast might prove to be not relevant in native conditions.

Conclusion:

AT4G32250 interacts with Toc159 and an exciting perspective would be to assess the ability of AT4g32250 to phosphorylate the Toc159 A-domain either *in vivo* as it was shown for SnrK2 (Pengcheng Wang et al. 2013) or with an *in vitro* assay on a recombinant A-domain as was done with the recombinant maize CK2 (Agne et al. 2010).

Tic56 was shown to co-precipitate with Toc159 even though it was originally purified in a complex containing only Tic20, Tic100 and Tic214 (Shingo Kikuchi et al. 2013). As already mentioned, Köhler et al. (2015) showed that *tic56* still imports a large spectrum of proteins suggesting that the 1MDa Tic complex is likely redundant with another Tic import machinery. A proteomic approach with Tic110, one of the major alternatives to the general 1MDa Tic complex might be interesting. As *tic110* homozygous mutant is embryo-lethal, an inducible silencing line might be used.

It is difficult to propose a function for Emb2004 as its embryo lethal phenotype might be the consequence of mutations in a variety of important chloroplastic functions. However, three functions are particularly represented in embryo lethal mutants of chloroplastic protein genes, namely protein synthesis, RNA maturation and processing and biosynthesis of amino acids, fatty acids and nucleotides (Bryant et al. 2011). Beside protein import, mutations that induce pale phenotypes affect particularly RNA synthesis, protein translation and folding, and isoprenoid synthesis (Gutiérrez-Nava et al. 2004). Among the three putative Emb2004 interaction partners identified in the yeast two hybrids screen, AT5G57710 / SMAX1 might be the more interesting as it is similar to chaperones, which act downstream of protein synthesis by helping newly synthetized protein to fold. There are approx. 200 LRR receptor kinases in plants. They are involved in plant-pathogen interactions, cell proliferation, hormone perception and can bind a variety of non-proteic ligand. The LRR domain acts as a receptor at the external side of the plasma membrane which activates a cytosolic kinase

upon ligand binding. The distance between the transit peptide cleavage site and the predicted transmembrane domain in Emb2004 sequence is very short and no kinase domain is predicted at the N-terminal of the sequence. However, some of these above described receptors lack a kinase domain (Torii 2004). A function in signaling of small non-proteic ligand would be consistent with the apparent difficulty to isolate Emb2004 interacting partner *in planta*e. The confirmation of an interaction between Emb2004 and the proteins identified in the yeast two hybrid screen *in planta* and testing an eventual role of Emb2004 in signaling of small molecule could represent the basis for future attempts to characterize Emb2004 function.

References:

Agne, B. et al., 2009. A Toc159 import receptor mutant, defective in hydrolysis of GTP, supports preprotein import into chloroplasts. *Journal of Biological Chemistry*, 284(13), pp.8670–8679.

Agne, B. et al., 2009. A toc159 import receptor mutant, defective in hydrolysis of GTP, supports preprotein import into chloroplasts. *J Biol Chem*, 284(13), pp.8670–8679. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19188370.

Agne, B. et al., 2010. The acidic A-domain of Arabidopsis TOC159 occurs as a hyperphosphorylated protein. *Plant physiology*, 153(3), pp.1016–1030.

Albiniak, A.M., Baglieri, J. & Robinson, C., 2012. Targeting of lumenal proteins across the thylakoid membrane. *Journal of Experimental Botany*, 63(4), pp.1689–1698.

Armbruster, U. et al., 2009. Chloroplast proteins without cleavable transit peptides: Rare exceptions or a major constituent of the chloroplast proteome? *Molecular Plant*, 2(6), pp.1325–1335.

Arnold, K. et al., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22(2), pp.195–201. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16301204.

Aronsson, H. et al., 2006. In vivo assessment of the significance of phosphorylation of the Arabidopsis chloroplast protein import receptor, atToc33. *FEBS Letters*, 580(2), pp.649–655.

Baerenfaller, K. et al., 2008. Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. *Science (New York, N.Y.)*, 320(5878), pp.938–41. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=4&SID=Q1tekooTk4CamRSw3Av&page=1&doc=4 [Accessed April 9, 2015].

Baldwin, A. et al., 2005. A molecular-genetic study of the Arabidopsis Toc75 gene family. *Plant physiology*, 138(2), pp.715–733.

Bauer, J. et al., 2002. Essential role of the G-domain in targeting of the protein import receptor atToc159 to the chloroplast outer membrane. *Journal of Cell Biology*, 159(5), pp.845–854.

Bauer, J. et al., 2002. Essential role of the G-domain in targeting of the protein import receptor atToc159 to the chloroplast outer membrane. *J Cell Biol*, 159(5), pp.845–854. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12460988.

Bauer, J. et al., 2000. The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature*, 403(6766), pp.203–207.

Baulcombe, D., 2004. RNA silencing in plants. *Nature*, 431(7006), pp.356–63. Available at: http://dx.doi.org/10.1038/nature02874 [Accessed May 19, 2015].

Becker, T. et al., 2004. Preprotein recognition by the Toc complex. *The EMBO journal*, 23(3), pp.520–530.

Bernoux, M. et al., 2011. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell host & microbe*, 9(3), pp.200–11. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=9&SID=Y2GiLF9GZpGFRcNVjod&page=1&doc=1 [Accessed May 19, 2015].

Bertsch, U. et al., 1992. Identification, characterization, and DNA sequence of a functional "double" groES-like chaperonin from chloroplasts of higher plants. *Proceedings of the National Academy of Sciences*, 89(18), pp.8696–8700. Available at: http://www.pnas.org/content/89/18/8696.abstract?ijkey=412a4adb1130fbc43cf354b835bb0cf3df51ba87&keytype2=tf_ipsecsha [Accessed May 16, 2015].

Birnboim, H.C. & Doly, J., 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6), pp.1513–1523.

Bischof, S. et al., 2011. Plastid Proteome Assembly without Toc159: Photosynthetic Protein Import and Accumulation of N-Acetylated Plastid Precursor Proteins. *The Plant Cell*, 23(11), pp.3911–3928.

Blobel, G., 2000. Protein targeting (Nobel lecture). *Chembiochem : a European journal of chemical biology*, 1(2), pp.86–102. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11828402 [Accessed May 16, 2015].

Blobel, G. & Dobberstein, B., 1975. OF PROTEINS ACROSS MEMBRANES Nascent Immunoglobulin Light Chains. , 67, pp.835–851.

Blobel, G. & Dobberstein, B., 1975a. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *The Journal of cell biology*, 67(3), pp.835–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2111658&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Blobel, G. & Dobberstein, B., 1975b. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *The Journal of cell biology*, 67(3), pp.852–62. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2111655&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Blobel, G. & Sabatini, D., 1971. Dissociation of Mammalian Polyribosomes into Subunits by Puromycin. *Proceedings of the National Academy of Sciences*, 68(2), pp.390–394. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=4&SID=T2tleqn89eTwTnJ4CQn&page=1&doc=9 [Accessed May 20, 2015].

Blom, N. et al., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6), pp.1633–1649.

Bob Buchanan, Wilhelm Gruissem, R.J., 2002. *Biochemistry & Molecular Biology of Plants*, John Wiley and Sons.

Bölter, B., May, T. & Soll, J., 1998. A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Letters*, 441(1), pp.59–62.

Boudreau, E. et al., 1997. A large open reading frame (orf1995) in the chloroplast DNA of Chlamydomonas reinhardtii encodes an essential protein. *MOLECULAR & GENERAL GENETICS*, 253(5), pp.649–653. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=130&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=7 [Accessed May 16, 2015].

Bourne, H.R., Sanders, D.A. & McCormick, F., 1990. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature*, 348(6297), pp.125–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2122258 [Accessed May 17, 2015].

Bruce, B.D., 2000. Chloroplast transit peptides: Structure, function and evolution. *Trends in Cell Biology*, 10(10), pp.440–447.

Bryant, N. et al., 2011. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol*, 155(4), pp.1678–1689. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21139083.

Buck, M., 1998. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quarterly reviews of biophysics*, 31(3), pp.297–355. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10384688 [Accessed May 16, 2015].

Carrie, C., Giraud, E. & Whelan, J., 2009. Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *The FEBS journal*, 276(5), pp.1187–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19187233 [Accessed May 16, 2015].

Chen, D. & Schnell, D.J., 1997. Insertion of the 34-kDa chloroplast protein import component, IAP34, into the chloroplast outer membrane is dependent on its intrinsic GTP- binding capacity. *Journal of Biological Chemistry*, 272(10), pp.6614–6620.

Chen, K.Y. & Li, H.M., 2007. Precursor binding to an 880-kDa Toc complex as an early step during active import of protein into chloroplasts. *Plant Journal*, 49(1), pp.149–158.

Chen, X., 2002. In Vivo Analysis of the Role of atTic20 in Protein Import into Chloroplasts. *THE PLANT CELL ONLINE*, 14(3), pp.641–654. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=95&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Chen, Y.-T. et al., 2013. ABA and the ubiquitin E3 ligase KEEP ON GOING affect proteolysis of the Arabidopsis thaliana transcription factors ABF1 and ABF3. *The Plant journal : for cell and molecular biology*, 75(6), pp.965–76. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3823012&tool=pmcentrez&rendertype=abstract [Accessed April 19, 2015].

Cherfils, J. & Zeghouf, M., 2013. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiological reviews*, 93(1), pp.269–309. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23303910 [Accessed January 20, 2015].

Chiu, C.-C. & Li, H.-M., 2008. Tic40 is important for reinsertion of proteins from the chloroplast stroma into the inner membrane. *The Plant journal : for cell and molecular biology*, 56(5), pp.793–801. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=91&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=4 [Accessed May 16, 2015].

Chou, M. et al., 2003. in the Chloroplast Protein Translocon. , 22(12).

Chou, M.L. et al., 2006. Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. *Journal of Cell Biology*, 175(6), pp.893–900.

Chua, N.H. & Schmidt, G.W., 1978. Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase. *Proceedings of the National Academy of Sciences of the United States of America*, 75(12), pp.6110–6114.

Cline, K. et al., 1985. Precursors to two nuclear-encoded chloroplast proteins bind to the outer envelope membrane before being imported into chloroplasts. *Journal of Biological Chemistry*, 260(6), pp.3691–3696.

Constan, D. et al., 2004. An outer envelope membrane component of the plastid protein import apparatus plays an essential role in Arabidopsis. *The Plant journal : for cell and molecular biology*, 38(1), pp.93–106. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15053763 [Accessed May 16, 2015].

Dahlin, C. & Cline, K., 1991. Developmental Regulation of the Plastid Protein Import Apparatus. *The Plant cell*, 3(10), pp.1131–1140.

Drescher, A. et al., 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant Journal*, 22(2), pp.97–104.

Duby, G., Oufattole, M. & Boutry, M., 2003. Hydrophobic residues within the predicted N-terminal amphiphilic α-helix of a plant mitochondrial targeting presequence play a major role in in vivo import. *The Plant Journal*, 27(6), pp.539–549. Available at: http://doi.wiley.com/10.1046/j.1365-313X.2001.01098.x [Accessed May 17, 2015].

Dutta, S., Mohanty, S. & Tripathy, B.C., 2009. Role of temperature stress on chloroplast biogenesis and protein import in pea. *Plant physiology*, 150(2), pp.1050–61. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2689951&tool=pmcentrez&rendertype=abstract [Accessed April 28, 2015].

Dutta, S., Teresinski, H.J. & Smith, M.D., 2014. A Split-Ubiquitin Yeast Two-Hybrid Screen to Examine the Substrate Specificity of atToc159 and atToc132, Two Arabidopsis Chloroplast Preprotein Import Receptors. *PloS one*, 9(4), p.e95026. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3988174&tool=pmcentrez&rendertype=abstract.

Emanuelsson, O. et al., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols*, 2(4), pp.953–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17446895 [Accessed July 11, 2014].

Emanuelsson, O., Nielsen, H. & Heijne, G.V.O.N., 1999. ChloroP , a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. , pp.978–984.

Feller, U., Anders, I. & Mae, T., 2008. Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *Journal of experimental botany*, 59(7), pp.1615–24. Available at: http://jxb.oxfordjournals.org/content/59/7/1615 [Accessed May 16, 2015].

Ferro, M. et al., 2010. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Molecular & cellular proteomics : MCP*, 9(6), pp.1063–84. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2877971&tool=pmcentrez&rendertype=abstract [Accessed May 19, 2015].

Ferro, M. et al., 2003. Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Molecular & cellular proteomics : MCP*, 2(5), pp.325–345.

Fink, A.L., 1999. Chaperone-Mediated Protein Folding. *Physiol Rev*, 79(2), pp.425–449. Available at: http://physrev.physiology.org/content/79/2/425 [Accessed May 16, 2015].

Flores-Pérez, Ú. & Jarvis, P., 2012. Molecular chaperone involvement in chloroplast protein import. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(2), pp.332–340. Available at: http://dx.doi.org/10.1016/j.bbamcr.2012.03.019.

Forsthoefel, N.R. et al., 2005. PIRLs: A novel class of plant intracellular leucine-rich repeat proteins. *Plant and Cell Physiology*, 46(6), pp.913–922.

Friedman, a L. & Keegstra, K., 1989. Chloroplast protein import : quantitative analysis of precursor binding. *Plant physiology*, 89(3), pp.993–999.

Froehlich, J., 2011. Studying Arabidopsis envelope protein localization and topology using thermolysin and trypsin proteases. *Methods in molecular biology (Clifton, N.J.)*, 774, pp.351–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21822849 [Accessed May 19, 2015].

Froehlich, J.E. et al., Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *Journal of proteome research*, 2(4), pp.413–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12938931 [Accessed May 19, 2015].

Glaser, S. et al., 2012. Tic22 is an essential chaperone required for protein import into the apicoplast. *Journal of Biological Chemistry*, 287(47), pp.39505–39512.

Gutiérrez-Nava, M.D.L.L. et al., 2004. CHLOROPLAST BIOGENESIS genes act cell and noncell autonomously in early chloroplast development. *Plant physiology*, 135(1), pp.471–482.

Harikrishna, K., Darby, R. & Draper, J., 1992. Chloroplast Dedifferentiation in Mechanically Isolated Asparagus Cells during Culture Initiation. *Plant physiology*, 100(3), pp.1177–83. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1075763&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Heazlewood, J.L. et al., 2007. SUBA: the Arabidopsis Subcellular Database. *Nucleic acids research*, 35(Database issue), pp.D213–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1635339&tool=pmcentrez&rendertype=abstract [Accessed June 24, 2015].

Von Heijne, G. & Nishikawa, K., 1991. Chloroplast transit peptides the perfect random coil? *FEBS Letters*, 278(1), pp.1–3. Available at: http://www.sciencedirect.com/science/article/pii/001457939180069F [Accessed May 16, 2015].

Heins, L. et al., 2002. The preprotein conducting channel at the inner envelope membrane of plastids. *The EMBO journal*, 21(11), pp.2616–25. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=80&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=3 [Accessed May 16, 2015].

Hernández Torres, J., Maldonado, M.A.A. & Chomilier, J., 2007. Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. *Biochemical and biophysical research communications*, 364(2), pp.325–31. Available at: https://hal-univ-diderot.archives-ouvertes.fr/hal-00195560/ [Accessed May 16, 2015].

Hernández-Torres, J., Jaimes-Becerra, A. & Chomilier, J., 2014. Arabidopsis thaliana Tic110, involved in chloroplast protein translocation, contains at least fourteen highly divergent heat-like repeated motifs. *Biologia*, 69(2), pp.139–151. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=75&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed March 25, 2015].

Hiltbrunner, A. et al., 2004. AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery. *Plant Molecular Biology*, 54(3), pp.427–440.

Hiltbrunner, A. et al., 2001. Targeting of an abundant cytosolic form of the protein import receptor at Toc159 to the outer chloroplast membrane. *Journal of Cell Biology*, 154(2), pp.309–316.

Hiltbrunner, A., 2001. Targeting of an abundant cytosolic form of the protein import receptor at Toc159 to the outer chloroplast membrane. *J Cell Biol*, 154(2), pp.309–316.

Hinnah, S. et al., 2002. The chloroplast protein import channel Toc75: Pore properties and interaction with transit peptides. *BIOPHYSICAL JOURNAL*, 83(2), pp.899–911. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=69&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=6 [Accessed May 16, 2015].

Hinnah, S.C. et al., 1997. Reconstitution of a chloroplast protein import channel. *The EMBO journal*, 16(24), pp.7351–60. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=59&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=2 [Accessed May 16, 2015].

Hirabayashi, Y. et al., 2011. In vivo studies on the roles of two closely related Arabidopsis Tic20 proteins, AtTic20-I and AtTic20-IV. *Plant & cell physiology*, 52(3), pp.469–78. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=93&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=2 [Accessed May 16, 2015].

Hofman, K. & Stoffel, W., TMBASE: a database of membrane spanning protein segments. Available at: http://www.researchgate.net/publication/248105554_TMBASE_a_database_of_membrane_spanning_protein_segments [Accessed May 19, 2015].

Hofmann, N.R. & Theg, S.M., 2005. Chloroplast outer membrane protein targeting and insertion. *Trends in Plant Science*, 10(9), pp.450–457.

Hooper, C.M. et al., 2014. Data and Text Mining SUBAcon : A consensus algorithm for unifying the subcellular lo- calization data of the Arabidopsis proteome. , 30(23), pp.1–8.

Inaba, T. et al., 2005. Arabidopsis tic110 is essential for the assembly and function of the protein import machinery of plastids. *The Plant cell*, 17(5), pp.1482–1496.

Inaba, T. et al., 2003. atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *Journal of Biological Chemistry*, 278(40), pp.38617–38627.

Infanger, S. et al., 2011. The chloroplast import receptor Toc90 partially restores the accumulation of Toc159 client proteins in the Arabidopsis thaliana ppi2 mutant. *Mol Plant*, 4(2), pp.252–263. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21220583.

Inoue, H., Li, M. & Schnell, D.J., 2013. An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), pp.3173–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3581895&tool=pmcentrez&rendertype=abstract.

Inoue, H., Rounds, C. & Schnell, D.J., 2010. The molecular basis for distinct pathways for protein import into Arabidopsis chloroplasts. *Plant Cell*, 22(6), pp.1947–1960. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20562235.

Inoue, K. et al., 2001. The N-terminal portion of the preToc75 transit peptide interacts with membrane lipids and inhibits binding and import of precursor proteins into isolated chloroplasts. *European Journal of Biochemistry*, 268(14), pp.4036–4043. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=72&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=2 [Accessed May 16, 2015].

Ivanova, Y. et al., 2004. Members of the Toc159 Import Receptor Family Represent Distinct Pathways for Protein Targeting to Plastids. , 15(July), pp.3379–3392.

Jackson, D.T., Froehlich, J.E. & Keegstra, K., 1998. The hydrophilic domain of tic110, and inner envelope membrane component of the chloroplastic protein translocation apparatus, faces stromal compartment. *Journal of Biological Chemistry*, 273(26), pp.16583–16588.

Jackson-Constan, D. & Keegstra, K., 2001. Arabidopsis genes encoding components of the chloroplastic protein import apparatus. *Plant physiology*, 125(4), pp.1567–1576.

Jarvis, P. et al., 1998. An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science (New York, N.Y.)*, 282(5386), pp.100–103.

Jarvis, P. & López-Juez, E., 2013. Biogenesis and homeostasis of chloroplasts and other plastids. *Nature reviews. Molecular cell biology*, 14, pp.787–802. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24263360.

Jarvis, P. & Soll, J., 2001. Toc, Tic, and chloroplast protein import. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1541(1-2), pp.64–79.

Jeon, Y. et al., 2010. Silencing of NbCEP1 encoding a chloroplast envelope protein containing 15 leucine-rich-repeats disrupts chloroplast biogenesis in Nicotiana benthamiana. *Mol Cells*, 29(2), pp.175–183. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20016945.

Jin, J.B. et al., 2001. A new dynamin-like protein, ADL6, is involved in trafficking from the trans-Golgi network to the central vacuole in Arabidopsis. *The Plant cell*, 13(7), pp.1511–1526.

Johnson, M. et al., 2008. NCBI BLAST: a better web interface. *Nucleic acids research*, 36(Web Server issue), pp.W5–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2447716&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2015].

Johnson, N., Powis, K. & High, S., 2013. Post-translational translocation into the endoplasmic reticulum. *Biochimica et biophysica acta*, 1833(11), pp.2403–9. Available at: http://www.sciencedirect.com/science/article/pii/S0167488912003862 [Accessed April 26, 2015].

Karim, S. & Aronsson, H., 2014. The puzzle of chloroplast vesicle transport - involvement of GTPases. *Frontiers in plant science*, 5, p.472. Available at: http://journal.frontiersin.org/article/10.3389/fpls.2014.00472/abstract [Accessed May 20, 2015].

Kasmati, A.R. et al., 2011. Molecular and genetic analyses of Tic20 homologues in Arabidopsis thaliana chloroplasts. *Plant J*, 66(5), pp.877–889. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21395885.

Keegstra, K., Olsen, L.J. & Theg, S.M., 1989. Chloroplastic Precursors and their Transport Across the Envelope Membranes. *Annual Review of Plant Physiology and Plant Molecular Biology*, 40(1), pp.471–501. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.pp.40.060189.002351 [Accessed May 16, 2015].

Kessler, F. et al., 1994. Identification of two GTP-binding proteins in the chloroplast protein import machinery. *Science*, 266(5187), pp.1035–1039. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7973656.

Kessler, F. & Blobel, G., 1996. Interaction of the protein import and folding machineries of the chloroplast. *Proceedings of the National Academy of Sciences of the United States of America*, 93(15), pp.7684–7689.

Kikuchi, S. et al., 2009. A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. *Plant Cell*, 21(6), pp.1781–1797. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19531596.

Kikuchi, S. et al., 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science*, 339(6119), pp.571–574. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23372012.

Kikuchi, S. et al., 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science (New York, N.Y.)*, 339(6119), pp.571–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23372012.

Kikuchi, S., Hirohashi, T. & Nakai, M., 2006. Characterization of the preprotein translocon at the outer envelope membrane of chloroplasts by blue native PAGE. *Plant Cell Physiol*, 47(3), pp.363–371. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16415065.

Köhler, D. et al., 2015a. Characterization of chloroplast protein import without Tic56, a component of the 1-MDa TIC translocon. *Plant Physiology*, 167(March), p.pp.114.255562. Available at: http://www.plantphysiol.org/lookup/doi/10.1104/pp.114.255562.

Köhler, D. et al., 2015b. Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts. *Plant physiology*, 167(3), pp.972–90. Available at: http://www.plantphysiol.org/content/early/2015/01/14/pp.114.255562 [Accessed April 17, 2015].

Kouranov, A. et al., 1998. Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *Journal of Cell Biology*, 143(4), pp.991–1002.

Kouranov, A. & Schnell, D.J., 1997. Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. *Journal of Cell Biology*, 139(7), pp.1677–1685.

Kovacheva, S. et al., 2005. In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. *Plant Journal*, 41(3), pp.412–428.

Kovacs-Bogdan, E. et al., 2011. Tic20 forms a channel independent of Tic110 in chloroplasts. *BMC Plant Biol*, 11, p.133. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21961525.

Kovács-Bogdán, E., Soll, J. & Bölter, B., 2010. Protein import into chloroplasts: The Tic complex and its regulation. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1803(6), pp.740–747. Available at: http://dx.doi.org/10.1016/j.bbamcr.2010.01.015.

Krimm, I. et al., 1999. A coil-helix instead of a helix-coil motif can be induced in a chloroplast transit peptide from Chlamydomonas reinhardtii. *European journal of biochemistry / FEBS*, 265(1), pp.171–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10491171 [Accessed May 16, 2015].

Kubis, S. et al., 2004. Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *Plant Cell*, 16(8), pp.2059–2077. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15273297.

Kubis, S. et al., 2003. The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *The Plant cell*, 15(8), pp.1859–1871.

De la Fuente van Bentem, S. et al., 2008. Site-specific phosphorylation profiling of Arabidopsis proteins by mass spectrometry and peptide chip analysis. *Journal of proteome research*, 7(6), pp.2458–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18433157 [Accessed May 17, 2015].

Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17846036 [Accessed July 9, 2014].

Leheny, E.A. & Theg, S.M., 1994. Apparent Inhibition of Chloroplast Protein Import by Cold Temperatures Is Due to Energetic Considerations Not Membrane Fluidity. *The Plant cell*, 6(3), pp.427–437. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=160445&tool=pmcentrez&rendertype=abstract [Accessed May 17, 2015].

Leipe, D.D. et al., 2002. Classification and evolution of P-loop GTPases and related ATPases. *Journal of molecular biology*, 317(1), pp.41–72.

Li, H.M. & Chen, L.J., 1997. A novel chloroplastic outer membrane-targeting signal that functions at both termini of passenger polypeptides. *Journal of Biological Chemistry*, 272(16), pp.10968–10974.

Li, H.M. & Teng, Y.S., 2013. Transit peptide design and plastid import regulation. *Trends in Plant Science*, 18(7), pp.360–366. Available at: http://dx.doi.org/10.1016/j.tplants.2013.04.003.

Ling, Q. et al., 2012. Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science*, 338(6107), pp.655–659. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23118188.

Lippold, F. et al., 2012. Fatty acid phytyl ester synthesis in chloroplasts of Arabidopsis. *The Plant cell*, 24(5), pp.2001–14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3442583&tool=pmcentrez&rendertype=abstract [Accessed September 17, 2014].

Lu, P. et al., 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology*, 25(1), pp.117–124.

Lubben, T.H., Theg, S.M. & Keegstra, K., 1988. Transport of proteins into chloroplasts. *Photosynthesis research*, 17(1-2), pp.173–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24429668 [Accessed May 16, 2015].

Ma, L. et al., 2001. Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *The Plant cell*, 13(12), pp.2589–607. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139475&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Ma, Y. et al., 1996. Two components of the chloroplast protein import apparatus, IAP86 and IAP75, interact with the transit sequence during the recognition and translocation of precursor proteins at the outer envelope. *Journal of Cell Biology*, 134(2), pp.315–327.

Madueno, F., Napier, J. & Gray, J., 1993. Newly Imported Rieske Iron-Sulfur Protein Associates with Both Cpn60 and Hsp70 in the Chloroplast Stroma. *The Plant cell*, 5(12), pp.1865–1876.

May, T., 2000. 14-3-3 Proteins Form a Guidance Complex with Chloroplast Precursor Proteins in Plants. *THE PLANT CELL ONLINE*, 12(1), pp.53–64. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=99&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

McFadden, G., 1999. Endosymbiosis and evolution of the plant cell. *Curr Opin Plant Biol*, 2(6), pp.513–519. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=10607659\npapers2://publication/uuid/58C0B1F6-AF33-4C75-AA53-4F3007AF56E9.

Milstein C, Brownlee GG, Harrison TM, M.M., 1972. A possible precurser of immunoglobulin light chains. *Nat New Biol.*, 239, pp.117–20.

Miras, S. et al., 2007. Toc159- and Toc75-independent import of a transit sequence-less precursor into the inner envelope of chloroplasts. *Journal of Biological Chemistry*, 282(40), pp.29482–29492.

Mishkind, M.L., Wessler, S.R. & Schmidt, G.W., 1985. Functional determinants in transit sequences: import and partial maturation by vascular plant chloroplasts of the ribulose-1,5-bisphosphate carboxylase small subunit of Chlamydomonas. *The Journal of cell biology*, 100(1), pp.226–34. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2113491&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Mitchell, A. et al., 2014. The InterPro protein families database: the classification resource after 15 years. *Nucleic acids research*, 43(D1), pp.D213–221. Available at: http://nar.oxfordjournals.org/content/43/D1/D213 [Accessed November 28, 2014].

Nielsen, E. et al., 1997. Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO Journal*, 16(5), pp.935–946.

Olsen, L.J. & Keegstra, K., 1992. The binding of precursor proteins to chloroplasts requires nucleoside triphosphates in the intermembrane space. *Journal of Biological Chemistry*, 267(1), pp.433–439.

Oreb, M. et al., 2008. Phosphorylation regulates the assembly of chloroplast import machinery. *Journal of experimental botany*, 59(9), pp.2309–16. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=28&SID=S1QNhy4osEOzRfw3fPs&page=1&doc=1 [Accessed May 17, 2015].

Oreb, M. et al., 2011. Substrate binding disrupts dimerization and induces nucleotide exchange of the chloroplast GTPase Toc33. *Biochem J*, 436(2), pp.313–319. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21434866.

Paila, Y.D., Richardson, L.G.L. & Schnell, D.J., 2015. New Insights into the Mechanism of Chloroplast Protein Import and Its Integration with Protein Quality Control, Organelle Biogenesis and Development. *Journal of Molecular Biology*, 427(5), pp.1038–1060. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0022283614004598.

Pain, D. et al., 1990. Identification and characterization of receptors for protein import into chloroplasts and mitochondria. *Indian journal of biochemistry & biophysics*, 27(6), pp.443–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1966338 [Accessed May 16, 2015].

Pay, A. et al., 2002. Plant RanGAPs are localized at the nuclear envelope in interphase and associated with microtubules in mitotic cells. *Plant Journal*, 30(6), pp.699–709.

Peoples, M.B. & Dalling, M.J., 1988. *Senescence and Aging in Plants*, Elsevier. Available at: http://www.sciencedirect.com/science/article/pii/B9780125209205500122 [Accessed May 16, 2015].

Perry, S.E. et al., 1991. Synthetic analogues of a transit peptide inhibit binding or translocation of chloroplastic precursor proteins. *Journal of Biological Chemistry*, 266(18), pp.11882–11889.

Perry, S.E. & Keegstra, K., 1994. Envelope membrane proteins that interact with chloroplastic precursor proteins. *The Plant cell*, 6(1), pp.93–105.

Pilon, M., de Kruijff, B. & Weisbeek, P.J., 1992. New insights into the import mechanism of the ferredoxin precursor into chloroplasts. *The Journal of biological chemistry*, 267(4), pp.2548–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1733952 [Accessed May 17, 2015].

Pokrovskaya, I.D. et al., 2011. Conserved oligomeric Golgi complex specifically regulates the maintenance of Golgi glycosylation machinery. *Glycobiology*, 21(12), pp.1554–1569.

Primavesi, L.F. et al., 2008. Visualisation of plastids in endosperm, pollen and roots of transgenic wheat expressing modified GFP fused to transit peptides from wheat SSU RubisCO, rice FtsZ and maize ferredoxin III proteins. *Transgenic research*, 17(4), pp.529–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17710559 [Accessed May 17, 2015].

Pyke, K. a. & Leech, R.M., 1994. A Genetic Analysis of Chloroplast Division and Expansion in Arabidopsis thaliana. *Plant physiology*, 104(1), pp.201–207.

Rahim, G. et al., 2009. In vivo interaction between atToc33 and atToc159 GTP-binding domains demonstrated in a plant split-ubiquitin system. *Journal of Experimental Botany*, 60(1), pp.257–267.

Ratnayake, R.M.U. et al., 2008. Alternative processing of Arabidopsis Hsp70 precursors during protein import into chloroplasts. *Bioscience, biotechnology, and biochemistry*, 72(11), pp.2926–35. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=103&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=2 [Accessed May 16, 2015].

Reiland, S. et al., 2009. Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant physiology*, 150(2), pp.889–903.

Rensink, W., Schnell, D. & Weisbeek, P., 2000. The transit sequence of ferredoxin contains different domains for translocation across the outer and inner membrane of the chloroplast envelope. , 275(14), pp.10265–10271. Available at: http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=148&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Rensink, W.A., 1998. Domains of a Transit Sequence Required for in Vivo Import in Arabidopsis Chloroplasts. *PLANT PHYSIOLOGY*, 118(2), pp.691–699. Available at: http://www.plantphysiol.org/content/118/2/691.long [Accessed May 16, 2015].

Reumann, S., Davila-Aponte, J. & Keegstra, K., 1999. The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: identification of a cyanobacterial homolog. *Proceedings of the National Academy of Sciences of the United States of America*, 96(2), pp.784–789.

Reumann, S. & Keegstra, K., 1999. The endosymbiotic origin of the protein import machinery of chloroplastic envelope membranes. *Trends in Plant Science*, 4(8), pp.302–307.

Rial, D. V, Ottado, J. & Ceccarelli, E.A., 2003. Precursors with altered affinity for Hsp70 in their transit peptides are efficiently imported into chloroplasts. *The Journal of biological chemistry*, 278(47), pp.46473–81. Available at: http://www.jbc.org/content/278/47/46473.long [Accessed May 16, 2015].

Richardson, L.G., Jelokhani-Niaraki, M. & Smith, M.D., 2009. The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. *BMC Biochem*, 10, p.35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20042108.

Richter, S. & Lamppa, G., 1999. Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. , 147(1), pp.33–43. Available at: http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=43&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Richter, S. & Lamppa, G.K., 2002. Determinants for removal and degradation of transit peptides of chloroplast precursor proteins. *The Journal of biological chemistry*, 277(46), pp.43888–94. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=44&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Robinson, C. & Bolhuis, A., 2004. Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp.135–147.

Rolland, N. et al., 2003. Proteomics of chloroplast envelope membranes. *Photosynthesis research*, 78(3), pp.205–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16245052 [Accessed May 19, 2015].

Rosano, G.L., Bruch, E.M. & Ceccarelli, E.A., 2011. Insights into the Clp/HSP100 chaperone system from chloroplasts of Arabidopsis thaliana. *The Journal of biological chemistry*, 286(34), pp.29671–80. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=105&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Row, P.E., 2001. Chloroplast precursor proteins compete to form early import intermediates in isolated pea chloroplasts. *Journal of Experimental Botany*, 52(354), pp.47–56. Available at: http://jxb.oxfordjournals.org/content/52/354/47.long [Accessed May 16, 2015].

Rudolf, M. et al., 2013. In vivo function of Tic22, a protein import component of the intermembrane space of chloroplasts. *Molecular Plant*, 6(3), pp.817–829.

Sagan, L., 1967. On the origin of mitosing cells. *Journal of theoretical biology*, 14(3), pp.255–274.

Sato, S. et al., 1999. Complete structure of the chloroplast genome of Arabidopsis thaliana. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 6, pp.283–290.

Schleiff, E., Jelic, M. & Soll, J., 2003. A GTP-driven motor moves proteins across the outer envelope of chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8), pp.4604–4609.

Schnell, D.J. et al., 1997. A consensus nomenclature for the protein-import components of the chloroplast envelope. *Trends in cell biology*, 7(8), pp.303–304.

Schnell, D.J. & Blobel, G., 1993. Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites. *Journal of Cell Biology*, 120(1), pp.103–116.

Schnell, D.J., Blobel, G. & Pain, D., 1991. Signal peptide analogs derived from two chloroplast precursors interact with the signal recognition system of the chloroplast envelope. *Journal of Biological Chemistry*, 266(5), pp.3335–3342.

Schnell, D.J., Kessler, F. & Blobel, G., 1994. Isolation of components of the chloroplast protein import machinery. *Science (New York, N.Y.)*, 266(5187), pp.1007–1012.

Schwartz, R.M. & Dayhoff, M.O., 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science (New York, N.Y.)*, 199(4327), pp.395–403. Available at: http://www.ncbi.nlm.nih.gov/pubmed/202030 [Accessed May 16, 2015].

Seedorf, M. & Soll, J., 1995. Copper chloride, an inhibitor of protein import into chloroplasts. *FEBS letters*, 367(1), pp.19–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7601278 [Accessed May 17, 2015].

Seedorf, M., Waegemann, K. & Soll, J., 1995. A constituent of the chloroplast import complex represents a new type of GTP-binding protein. *The Plant journal : for cell and molecular biology*, 7(3), pp.401–411.

Sigrist, C.J.A. et al., 2013. New and continuing developments at PROSITE. *Nucleic acids research*, 41(Database issue), pp.D344–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3531220&tool=pmcentrez&rendertype=abstract [Accessed December 21, 2014].

Smith, M.D. et al., 2004. atToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. *J Cell Biol*, 165(3), pp.323–334. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15138290.

Smith, M.D. et al., 2003. In vitro analysis of chloroplast protein import. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]*, Chapter 11, p.Unit11.16.

Smith, M.D. et al., 2002. The targeting of the atToc159 preprotein receptor to the chloroplast outer membrane is mediated by its GTPase domain and is regulated by GTP. *Journal of Cell Biology*, 159(5), pp.833–843.

Spreitzer, R.J. & Salvucci, M.E., 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annual review of plant biology*, 53, pp.449–475.

Stahl, T. et al., 1999. Tic40, a new "old" subunit of the chloroplast protein import translocon. *Journal of Biological Chemistry*, 274(52), pp.37467–37472.

Stengel, A. et al., 2009. Preprotein import into chloroplasts via the Toc and Tic complexes is regulated by redox signals in Pisum sativum. *Molecular plant*, 2(6), pp.1181–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19995724 [Accessed May 17, 2015].

Stengel, A. et al., 2008. TIC62 Redox-regulated Translocon Composition and Dynamics. *Journal of Biological Chemistry*, 283(11), pp.6656–6667. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18180301 [Accessed April 11, 2015].

Stephens, D.J. & Banting, G., 2000. The use of yeast two-hybrid screens in studies of protein:protein interactions involved in trafficking. *Traffic (Copenhagen, Denmark)*, 1(10), pp.763–768.

Su, P.-H. & Li, H., 2010. Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. *The Plant cell*, 22(5), pp.1516–1531.

Sun, Q. et al., 2009. PPDB, the Plant Proteomics Database at Cornell. *Nucleic acids research*, 37(Database issue), pp.D969–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686560&tool=pmcentrez&rendertype=abstract [Accessed May 18, 2015].

Sun, Y.-J. et al., 2002. Crystal structure of pea Toc34, a novel GTPase of the chloroplast protein translocon. *Nature structural biology*, 9(2), pp.95–100.

Sunilkumar, G. et al., 2002. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. *Plant Molecular Biology*, 50(3), pp.463–474.

Sveshnikova, N., Grimm, R., et al., 2000. Topology studies of the chloroplast protein import channel Toc75. *Biological chemistry*, 381(8), pp.687–93. Available at: http://www.degruyter.com/view/j/bchm.2000.381.issue-8/bc.2000.089/bc.2000.089.xml;jsessionid=E8B0839F64CE435ACF26829BC1E40A2F [Accessed May 16, 2015].

Sveshnikova, N., Soll, J. & Schleiff, E., 2000. Toc34 is a preprotein receptor regulated by GTP and phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, 97(9), pp.4973–4978.

Teng, Y.-S. et al., 2006. Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *The Plant cell*, 18(9), pp.2247–2257.

Teng, Y.-S., Chan, P.-T. & Li, H., 2012. Differential age-dependent import regulation by signal peptides. *PLoS biology*, 10(10), p.e1001416. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3484058&tool=pmcentrez&rendertype=abstract [Accessed May 17, 2015].

Thimm, O. et al., 2004. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, 37(6), pp.914–939.

Timmis, J.N. et al., 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature reviews. Genetics*, 5(2), pp.123–135.

Torii, K.U., 2004. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *International review of cytology*, 234, pp.1–46. Available at: http://www.sciencedirect.com/science/article/pii/S0074769604340015 [Accessed May 21, 2015].

Tranel, P.J. et al., 1995. A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway. *The EMBO journal*, 14(11), pp.2436–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=398357&tool=pmcentrez&rendertype=abstract [Accessed May 16, 2015].

Tranel, P.J. & Keegstra, K., 1996. A novel, bipartite transit peptide targets OEP75 to the outer membrane of the chloroplastic envelope. *The Plant cell*, 8(11), pp.2093–2104.

Tripp, J. et al., 2007. A novel serine/proline-rich domain in combination with a transmembrane domain is required for the insertion of AtTic40 into the inner envelope membrane of chloroplasts. *The Plant journal : for cell and molecular biology*, 52(5), pp.824–38. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=86&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=1 [Accessed May 16, 2015].

Trujillo, M. & Shirasu, K., 2010. Ubiquitination in plant immunity. *Current Opinion in Plant Biology*, 13(4), pp.402–408. Available at: http://dx.doi.org/10.1016/j.pbi.2010.04.002.

Tsai, J.Y. et al., 2013. Structural characterizations of the chloroplast translocon protein Tic110. *Plant Journal*, 75(5), pp.847–857.

Tsugeki, R. & Nishimura, M., 1993. Interaction of homologues of Hsp70 and Cpn60 with ferredoxin-NADP+ reductase upon its import into chloroplasts. *FEBS Letters*, 320(3), pp.198–202. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=120&SID=X24wXHSN7IsVY1Uh7mg&page=2&doc=18 [Accessed March 24, 2015].

Tzafrir, I. et al., 2004. Identification of genes required for embryo development in Arabidopsis. *Plant Physiol*, 135(3), pp.1206–1220. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15266054.

Villarejo, A. et al., 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nature cell biology*, 7(12), pp.1224–1231.

Waegemann, K. & Soil, J., 1991. Characterization of the protein import apparatus in isolated outer envelopes of chloroplasts. *The Plant Journal*, 1(2), pp.149–158. Available at: http://doi.wiley.com/10.1111/j.1365-313X.1991.00149.x [Accessed May 16, 2015].

Wan, J. et al., 1995. Import characteristics of a leucoplast pyruvate kinase are influenced by a 19-amino-acid domain within the protein. *The Journal of biological chemistry*, 270(28), pp.16731–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7622484 [Accessed May 17, 2015].

Wan, J. et al., 1996. Transit peptides play a major role in the preferential import of proteins into leucoplasts and chloroplasts. *The Journal of biological chemistry*, 271(49), pp.31227–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8940125 [Accessed May 17, 2015].

Wang, F. et al., 2008. The role of GTP binding and hydrolysis at the atToc159 preprotein receptor during protein import into chloroplasts. *Journal of Cell Biology*, 183(1), pp.87–99.

Wang, P. et al., 2013. Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. *Proc Natl Acad Sci U S A*, 110(27), pp.11205–11210. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23776212.

Wang, P. et al., 2013. Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), pp.11205–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3703982&tool=pmcentrez&rendertype=abstract.

Wardle, A. et al., 2007. Further in vivo studies on the role of the molecular chaperone , Hsp93 , in plastid protein import. , pp.364–379.

Waters, M.T. et al., 2014. The karrikin response system of Arabidopsis. *The Plant journal : for cell and molecular biology*, 79(4), pp.623–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24433542 [Accessed May 20, 2015].

Weibel, P. et al., 2003. Dimerization of Toc-GTPases at the chloroplast protein import machinery. *Journal of Biological Chemistry*, 278(39), pp.37321–37329.

Whiteman, S.-A. et al., 2008. Identification of novel proteins and phosphorylation sites in a tonoplast enriched membrane fraction of Arabidopsis thaliana. *Proteomics*, 8(17), pp.3536–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18686298 [Accessed May 17, 2015].

Wienk, H.L. et al., 2000. Structure, dynamics, and insertion of a chloroplast targeting peptide in mixed micelles. *Biochemistry*, 39(28), pp.8219–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10889029 [Accessed May 16, 2015].

Wienk, H.L.J., Czisch, M. & de Kruijff, B., 1999. The structural flexibility of the preferredoxin transit peptide. *FEBS Letters*, 453(3), pp.318–326. Available at: http://www.sciencedirect.com/science/article/pii/S0014579399006535 [Accessed May 16, 2015].

Wilson, D. et al., 2009. SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic acids research*, 37(Database issue), pp.D380–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686452&tool=pmcentrez&rendertype=abstract [Accessed April 13, 2015].

Yan, X. et al., 2006. Differential uptake of photosynthetic and non-photosynthetic proteins by pea root plastids. *FEBS letters*, 580(27), pp.6509–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17101134 [Accessed May 17, 2015].

Young, L., Cross, R. & Byun-McKay, S., 2005. A high- and low-temperature inducible Arabidopsis thaliana HSP101 promoter located in a nonautonomous Mutator-like element. , 48(3), pp.547–555. Available at: http://apps.isiknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=2&SID=T2tleqn89eTwTnJ4CQn&page=1&doc=1 [Accessed May 20, 2015].

Young, M.E., Keegstra, K. & Froehlich, J.E., 1999. GTP promotes the formation of early-import intermediates but is not required during the translocation step of protein import into chloroplasts. *Plant physiology*, 121(1), pp.237–244.

Zhang, C. & Guy, C.L., 2005. Co-immunoprecipitation of Hsp101 with cytosolic Hsc70. *Plant Physiology and Biochemistry*, 43(1), pp.13–18.

Zhong, R. et al., 2003. A pea antisense gene for the chloroplast stromal processing peptidase yields seedling lethals in Arabidopsis : survivors show defective GFP import in vivo. *The Plant Journal*, 34(6), pp.802–812. Available at: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=46&SID=X24wXHSN7IsVY1Uh7mg&page=1&doc=2 [Accessed May 16, 2015].

Zybailov, B. et al., 2008. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, 3.

Collaborations:

Mass-spectrometry:

The mass-spectrometry analysis of the TAP-Toc159 eluates was done by Sylvain Bischof in ETHZ, Zürich.

The mass-spectrometry analysis of the Emb2004-TAP eluate was done in Baginsky's Lab, Martin Luther Universität Halle-Wittenberg , Halle, Germany.

AT4G32250 project:

Mònica Arias Maldonado, a PhD student in Kessler's Lab is currently working on the AT4G32250 project, she kindly provided the anti-AT4G32250 antibody which she purified and information on the AT4G32250 phenotype.

The AT4G32250-YFP construct, the pET21d-AT4G32250 construct and the *in vitro* import/insertion experiment were done by Bastien Christ a former Master student in Kessler's Lab.

AT5G01590/Tic56 project:

The characterization of AT5G01590/Tic56 (microscopy, fig. 16) was done in Baginsky's Lab, Martin Luther Universität Halle-Wittenberg , Halle, Germany. My contribution to this project was the identification of AT5G01590/Tic56 in the TAP-Toc159 eluate (fig. 17).

Emb2004/At1G10510 project:

The trypsin protease protection assay was done in Baginsky's Lab, Martin Luther Universität Halle-Wittenberg.

Papers:

Contribution to published papers:

Agne et al. 2010: fig. 2, fig. 3, fig. 8.

Lippold et al. 2012: fig. 1.

Köhler et al. 2015: fig. 1

Acknowledgements:

I thank my thesis director P. Félix Kessler for giving me the opportunity to do my PhD in his lab, his advices during the project, and his revision of the manuscript of this thesis. I thank Dr. Birgit Agne (Martin Luther Universität Halle-Wittenberg , Halle, Germany) for her advices at the beginning of this thesis as she worked as a postdoc in Kessler's lab, and her later collaboration and advices as she moved in Germany. I thank P. Sacha Baginsky (Martin Luther Universität Halle-Wittenberg, Germany) for giving me the opportunity to visit his lab for two weeks, for the assistance of his bachelor students for the transient expression of Emb2004-YFP in *A. thaliana* protoplasts, the mass-spectrometry analysis and for the fruitful collaboration offered on the Tic56 project. Mònica Arias Maldonado is kindly acknowledged for the collaboration on the AT4G32250 project, for providing me the anti-At4G32250 antibody and her communication on the AT4G32250 homozygous mutant phenotype. Veronique Douet, scientific collaborator in Kessler's lab, is acknowledged for her help on many experiments, particularly the electron microscopy. I want to thank all the present and past collaborators of the LPV for their help, scientific inputs, good humor and kindness. Peoples from other labs, from the technical staff of the University and librarians made the life easier with their help, advices and with all sorts of technical supports when dealing with everyday problems. Finally I want to thank my family, especially my two parents, Fabienne and Paul-Etienne, who supported me on over the years.

Annex:

Tab. 1: Sequence of the identified by the in-house yeast two hybrid screen: The sequences amplified by PCR from the positive clones obtained by the in-house two hybrid screen were searched against the NCBI non-redundant nucleotides collection using the BLAST algorithm. Dark green: mRNA *A. thaliana*; green: *A. thaliana* rRNA; Light green: *A. thaliana* genomic sequence; Blue: Plant; Salmon: human; grey: yeast vector.

Clone:	Accession:	Description:	max score:	total score:	coverage:	E-value:	identity:
TM10	AK230153.1	Arabidopsis thaliana mRNA for hypothetical protein, complete cds, clone: RAFL22-85-G17	344	344	96%	2.00E-91	100%
TM14	AK230153.1	Arabidopsis thaliana mRNA for hypothetical protein, complete cds, clone: RAFL22-85-G17	597	597	77%	2.00E-167	99%
TM16	AK230153.1	Arabidopsis thaliana mRNA for hypothetical protein, complete cds, clone: RAFL22-85-G17	532	532	100%	3.00E-148	99%
TM17	XM_002863257.1	Arabidopsis lyrata subsp. lyrata ORF64c, mRNA	230	282	63%	7.00E-57	100%
TP21	XM_002863257.1	Arabidopsis lyrata subsp. lyrata ORF64c, mRNA	1293	1293	99%	0	100%
TM1	AP000423.1	Arabidopsis thaliana chloroplast DNA, complete genome, ecotype: Columbia	558	1117	75%	7.00E-156	100%
TM8	AP000423.1	Arabidopsis thaliana chloroplast DNA, complete genome, ecotype: Columbia	558	1117	98%	7.00E-156	100%
TM19	EZ733504.1	TSA: Arachis hypogaea CL1Contig12520.Arhy mRNA sequence	54.7	54.7	35%	4.00E-04	100%
TM22	EZ181330.1	TSA: Artemisia annua strain Artemis Contig4806.Ac mRNA sequence	193	193	99%	9.00E-46	100%
TM24	FN599860.1	Camelina sativa 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and partial 25S rRNA gene, cultivar Calena	311	311	96%	2.00E-81	100%
TM2	EZ733504.1	TSA: Arachis hypogaea CL1Contig12520.Arhy mRNA sequence	54.7	54.7	34%	4.00E-04	100%
TM3	FN599860.1	Camelina sativa 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and partial 25S rRNA gene, cultivar Calena	311	311	96%	2.00E-81	100%
TM4	FN599860.1	Camelina sativa 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and partial 25S rRNA gene, cultivar Calena	311	311	96%	2.00E-81	100%
TM6	EZ733504.1	TSA: Arachis hypogaea CL1Contig12520.Arhy mRNA sequence	62.1	171	64%	3.00E-06	100%
TM7	EZ728631.1	TSA: Arachis hypogaea CL1Contig7647.Arhy mRNA sequence	56.5	56.5	40%	1.00E-04	97%
TM13	EZ733504.1	TSA: Arachis hypogaea CL1Contig12520.Arhy mRNA sequence	54.7	54.7	36%	4.00E-04	100%
TM5	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TM9	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TM12	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	68%	0	100%
TM23	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	881	881	76%	0	100%
TP2	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	640	640	99%	3.00E-180	99%
TP6	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP7	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP8	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP9	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	640	640	99%	3.00E-180	99%
TP10	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	1421	1421	89%	0	99%
TP11	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP12	XR_093034.1	PREDICTED: Pongo abelii fibronectin-like (LOC100449704), miscRNA	117	117	30%	5.00E-23	98%

Clone:	Accession:	Description:	max score:	total score:	coverage:	E-value:	identity:
TP6	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP7	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP8	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP9	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	640	640	99%	3.00E-180	99%
TP10	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	1421	1421	89%	0	99%
TP11	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TP12	XR_093034.1	PREDICTED: Pongo abelii fibronectin-like (LOC100449704), miscRNA	117	117	30%	5.00E-23	98%
TP13	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	1509	1509	93%	0	98%
TP15	AB384664.2	Synthetic construct DNA, clone: pF1KB0006, Homo sapiens FN1 gene for fibronectin precursor, complete cds, without stop codon, in Flexi system	640	640	99%	3.00E-180	99%
TP22	EF550134.1	Homo sapiens fibronectin splice variant E (FN1) mRNA, partial cds, alternatively spliced	929	929	77%	0	100%
TM18	FJ696408.1	Yeast two-hybrid vector pGADCg, complete sequence	1286	1286	88%	0	99%