Structure-function analysis of the acidic domain of the Arabidopsis Toc159 receptors

by

Lynn Richardson

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2008

©Lynn Richardson 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Most chloroplast proteins are encoded in the nucleus and translated in the cytosol with an N-terminal transit peptide, which facilitates recognition by the receptors of the **t**ranslocon at the **o**uter membrane of **c**hloroplasts (Toc). The Toc159 family of receptors in *Arabidopsis thaliana* are the primary chloroplast preprotein receptors. Members of this family differentially associate with either atToc33 or atToc34 ("at" designates the species of origin, *Arabidopsis thaliana*) to form structurally and functionally distinct Toc complexes; atToc159/33-containing complexes import photosynthetic preproteins, and atToc132(120)/34-containing complexes import non-photosynthetic, plastid house-keeping proteins. The Toc159 receptors are most variable in their N-terminal A-domain, suggesting that this domain may contribute to their functional specificity. The A-domain has structural properties characteristic of intrinsically unstructured protein (IUP) domains, including an abundance of acidic amino acid residues, aberrant mobility during SDS-PAGE and sensitivity to proteolysis. The overall objective of this study was to gain insight into the function of the A-domain. First, to investigate the role of the A-domain in the assembly of structurally distinct Toc complexes, full-length, truncated and domain-swapped variants of atToc159 and atToc132 were targeted *in vitro* to chloroplasts isolated from wild type (WT) Arabidopsis, and atToc33 and atToc34 null mutants (*ppi*1 and *ppi*3, respectively). Insertion of atToc132 was less efficient than atToc159, and was not affected by the removal or swapping of the A-domain. In contrast, removal of the A-domain of atToc159 resulted in decreased insertion, most notably into *ppi*1 chloroplasts, suggesting that the A-domain is important for insertion, especially into atToc34-containing complexes. These results indicate that the A-domain does play a role in targeting, and may also suggest different roles for the A-domain in targeting of atToc159 and atToc132. Second, a structural analysis of the A-domain of atToc132 and atToc159 was performed using CD and fluorescence spectroscopy to gain insight into their potential function(s). The A-domains were found to be unstructured at physiological pH, and their secondary structure increased with increasing temperature and decreasing pH, which are characteristics of IUPs. IUPs are commonly involved in

protein-protein interactions, and their unstructured nature may suggest a role for the A-domains in binding transit peptides, accounting for the ability of the Toc159 receptors to differentially distinguish between a large number of diverse transit peptides that possess low sequence conservation.

Acknowledgements

I would first like to thank my supervisor, Dr. Matt Smith for his advice, encouragement and support (academic-related and otherwise), and for getting me interested in research as an undergraduate and graduate student – I've learned more in the past two years than I thought possible. I would also like to thank my co-supervisor Dr. Glick for his advice and for sharing his vast knowledge of all things molecular biology-related. Thank you to Dr. Masoud Jelokhani-Niaraki for his help with fluorescence and CD spectroscopy, for his very insightful discussions and sharing his passion for science. Thanks to Gena Braun for training me on various pieces of equipment, and for having a fantastic ability to just make things work. Thanks to Mike Pyne, Colin Yardley, Kyle Weston and Tony Straus for all of their help in the lab and to other past and present members of the Smith lab – Lisa Paige, Fern McSorley, Wes Farquharson, and Gabi Krnac – for the great chats and company, and making the lab a great place to work; and a special thanks to Marina Ivanova for her expertise in biochemistry (and allowing me to talk her into coffee breaks). I would also like to thank Michelle De Decker and Megan House for their thoughtful discussions and sharing in my lab frustrations. A very big thank you to my parents, Wilma and Steve Richardson for all of their support, and to Jill, Karis, Clint, Sandra and Josh, and other family and friends for reminding me that life outside of the lab *does* exist.

Table of Contents

LIST OF FIGURES VIII

LIST OF ABBREVIATIONS IX

1. INTRODUCTION 1

1.1 Plastid structure and function 1

1.2 Plastid Evolution 4

1.3 Chloroplast protein import 7

1.4 The general import apparatus 11
1.4.1 The Tic complex 11
1.4.2 The Toc complex 15

1.5 Multiple protein import pathways 16

1.6 The Arabidopsis Toc159 family of receptors 22
1.6.1 Targeting of Toc159 to chloroplasts during initial Toc complex assembly 22
1.6.2 Preprotein recognition by the Toc159 receptors 26

1.7 Overall objectives 27

2. INVESTIGATION OF THE ROLE OF THE A-DOMAIN IN TARGETING THE TOC159 RECEPTORS TO STRUCTURALLY DISTINCT TOC COMPLEXES 29

2.1 Background 29

2.2 Objectives and Hypotheses 29

2.3 Materials and Methods 30
2.3.1 Plant material and growth conditions 30
2.3.2 Isolation of intact chloroplasts from Arabidopsis thaliana 30
2.3.3 In vitro translation of radiolabelled A. thaliana Toc159 homologues 31
2.3.4 In vitro targeting assays 32
2.3.5 SDS-PAGE and phosphorimager analysis 33
2.3.6 Statistics 33

2.4 Results 35
2.4.1 Effect of the A-domain on targeting to wild-type chloroplasts 35
2.4.2 Effect of the A-domain on targeting to structurally distinct Toc complexes 43

2.5 Discussion 47
2.5.1 Effects of the A-domain on targeting of atToc159 and atToc132 to isolated chloroplasts 47
2.5.2 Role of the A-domain in targeting to structurally distinct Toc complexes 49
2.5.3 Summary 51

3. STRUCTURAL ANALYSIS OF THE TOC159 FAMILY A-DOMAINS 55

3.1 Background 55

3.2 Objectives and Hypotheses 56

3.3 Materials and methods 61
3.3.1 Generation of atTOC159A and atTOC132A constructs 61
3.3.2 Expression and purification of recombinant A-domains 62
3.3.3 Chemical modification of $132A_{His}$ and $159A_{His}$ 63
3.3.4 Western blots 64
3.3.5 Circular Dichroism Spectroscopy 65
3.3.6 Fluorescence Spectroscopy 65

3.4 Results 67
3.4.1 The A-domains of atToc159 and atToc132 are predicted to be natively disordered 67
3.4.2 Expression and purification of $132A_{His}$ and $159A_{His}$ 67
3.4.3 $132A_{His}$ and $159A_{His}$ are unstructured under physiological conditions 72
3.4.4 Temperature and pH dependent structural changes in $132A_{His}$ and $159A_{His}$ 77
3.4.5 Structure of $132A_{His}$ and $159A_{His}$ in the presence of TFE 82

3.5 Discussion 85
3.5.1 Potential role for the A-domain in the functional diversity within the Toc159 family 88

4. CONCLUSIONS 91

REFERENCES 93

APPENDICES 102

APPENDIX 1: *In vitro* targeting assay data 102

APPENDIX 2: Statistical analyses 103

APPENDIX 3: Primers used for PCR amplification of the A-domain of atToc159 and atToc132. 105

List of Figures

Figure 1. Chloroplast structure 3

Figure 2. Plastid evolution 6

Figure 3. Toc/Tic protein import pathway 9

Figure 4. Schematic diagram of the Toc and Tic complexes 13

Figure 5. Multiple protein import pathways in Arabidopsis 21

Figure 6. Amino acid sequence comparison of the Arabidopsis Toc159 receptors 24

Figure 7. Schematic representation of recombinant Toc159 receptor constructs used for *in vitro* targeting assays 37

Figure 8. Phosphorimager-visualized SDS-PAGE gels following in vitro targeting assays 39

Figure 9. Targeting efficiency of Toc159 receptor constructs to chloroplasts isolated from WT plants 41

Figure 10. Targeting efficiency of Toc159 receptor constructs to chloroplasts isolated from WT, *ppi1* and *ppi3* Arabidopsis 45

Figure 11. Representation of typical far-UV CD Spectra associated with different types of secondary structure 58

Figure 12. Fluorescence characteristics of aromatic amino acid residues 60

Figure 13. Disorder prediction of atToc132 and atToc159 69

Figure 14. SDS-PAGE and Western blots of $132A_{His}$ and $159A_{His}$ 71

Figure 15. EDC treatment of $132A_{His}$ and $159A_{His}$ 74

Figure 16. CD spectra of $132A_{His}$ (5 µM) and $159A_{His}$ (5 µM) at 25°C, pH 8 76

Figure 17. CD spectra of $132A_{His}$ and $159A_{His}$ with changing temperature 79

Figure 18. pH-dependent structural changes in $132A_{His}$ and $159A_{His}$ 81

Figure 19. CD spectra of $132A_{His}$ and $159A_{His}$ in Buffer/TFE mixtures 83

List of Abbreviations

132A159GM: construct of the A-domain of atToc132 fused to G- and M-domains of atToc159

132GM: truncated atToc132 construct, consisting of the G- and M-domains of atToc132

159A132GM: construct of the A-domain of atToc159 fused to G- and M-domains of atToc132

159GM: truncated atToc159 construct, consisting of the G- and M-domains of atToc159

AKR2A: ankyrin repeat protein 2A

CD: circular dichroism

ceQORH: chloroplast envelope quinone oxidoreductase

coGAP: co-GTPase activating protein

EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide

Hip/Hop domain: Hsc70-interacting protein/Hsc70/Hsp90-organizing protein domain

IMAC: immobilized metal affinity chromatography

IMS: intermembrane space

IUP: intrinsically unstructured protein

IVT: in vitro translation product

MGDG: monogalactosyldiacyl glycerol

NAD(P+): oxidized form of nicotinamide adenine dinucleotide phosphate

OEP: outer envelope protein

OMP: outer membrane protein

*ppi*1(2,3): plastid protein import mutant 1 (2,3)

pSSU: precursor of the small subunit of Rubisco

TFE: trifluoroethanol

Tic: translocon at the inner membrane of chloroplasts

Toc: translocon at the outer membrane of chloroplasts

TL: thermolysin

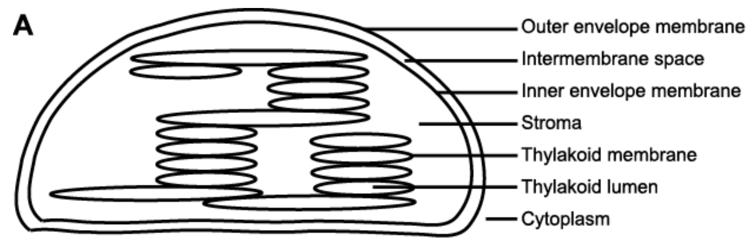
1. Introduction

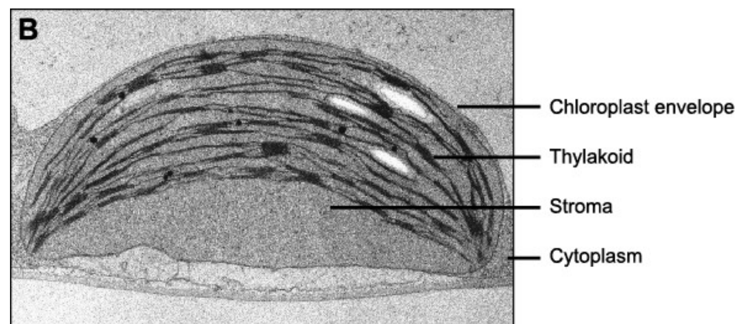
1.1 Plastid structure and function

Plastids are a structurally and functionally diverse group of organelles, indispensible for the growth and development of all plants. The most well-known plastid type is the chloroplast, the site of photosynthesis, as well as many other important biochemical reactions including fatty acid, lipid, amino acid and protein synthesis, as well as nitrogen and sulphur assimilation (Wise 2006). Other plastid types include proplastids, which are present in embryonic and meristematic tissue and are the precursor to all other plastid types; amyloplasts, starch storing and synthesizing plastids that are also involved in graviperception; elaioplasts, oil-storing plastids; chromoplasts, brightly coloured, carotenoid-storing plastids; and gerontoplasts, which arise from chloroplasts in senescing leaves (Wise 2006). The distinct functions of different plastid types may be largely attributed to their specific protein complements, as the biochemical pathways that make plastid types unique require specific sets of metabolic enzymes. While each type of plastid may perform different functions, plastid types are interconvertible, emphasizing the dynamic nature of this organelle.

All plastids are surrounded by a double envelope membrane; however internal structure varies between plastid types. Chloroplasts contain a highly structured internal thylakoid membrane system required for coordination of the molecular events of photosynthesis. This internal membrane system, along with the double membrane envelope separates the chloroplast into six distinct subcompartments; the outer membrane, intermembrane space, inner membrane, stroma, thylakoid membrane and thylakoid lumen (Figure 1; Smith and Schnell 2004). Other plastid types have a less well-structured internal membrane system relative to chloroplasts, but often contain other structures important for their function such as starch granules in amyloplasts, lipid bodies (called plastoglobuli) in elaioplasts, or carotenoid-containing plastoglobuli and fibrils in chromoplasts (Wise 2006).

Figure 1. Chloroplast structure. A) A diagram of a chloroplast outlining structural features and subcompartments. B) A transmission electron micrograph of a leaf chloroplast from *Arabidopsis thaliana* (Smith and Schnell, 2004).


A
Outer envelope membrane
Intermembrane space
Inner envelope membrane
Stroma
Thylakoid membrane
Thylakoid lumen
Cytoplasm

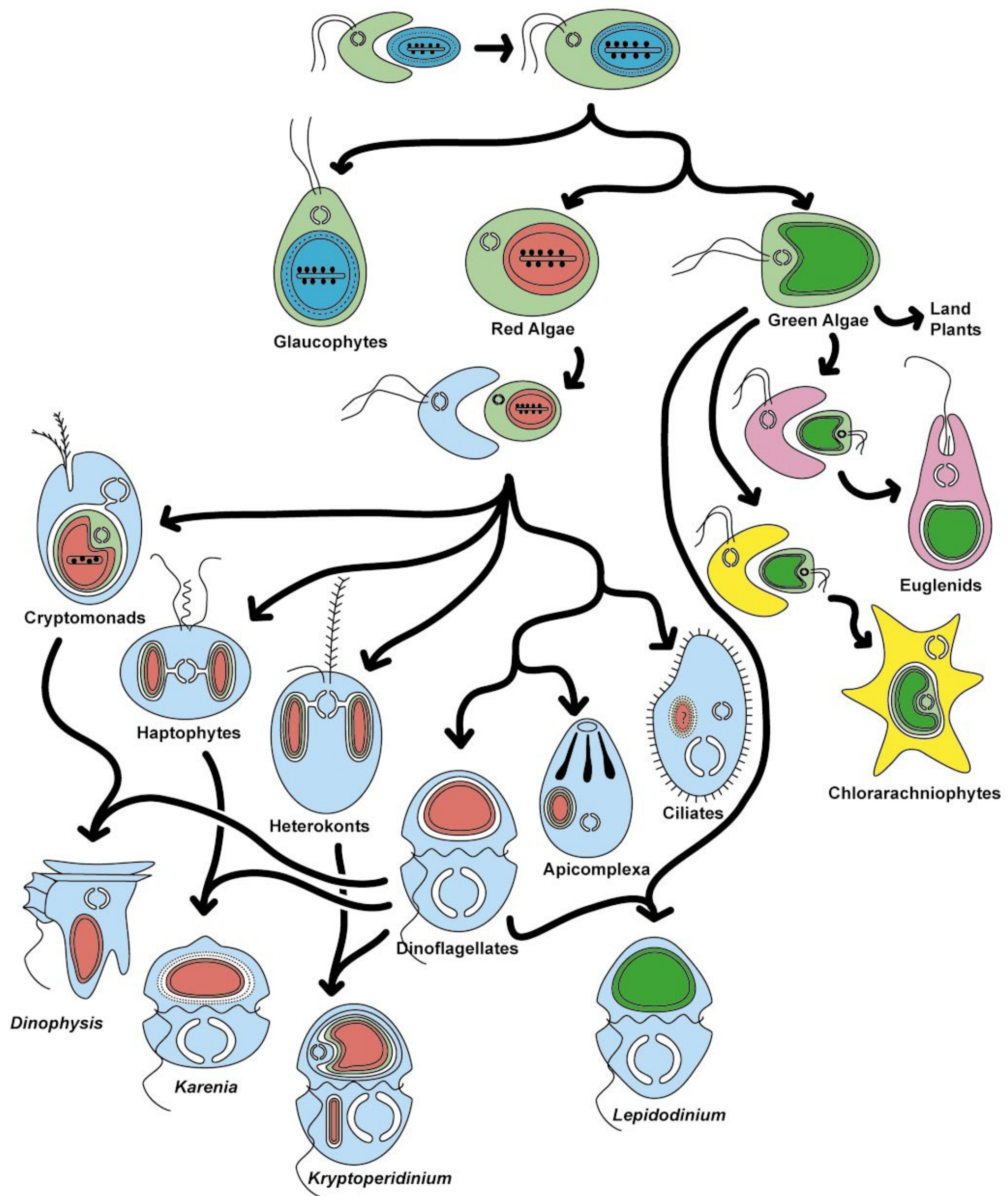

B
Chloroplast envelope
Thylakoid
Stroma
Cytoplasm

1.2 Plastid Evolution

The primary endosymbiotic event that led to the evolution of land plants is believed to have occurred approximately 1.2 to 1.5 billion years ago, from the uptake of a cyanobacterium by a mitochondria-containing eukaryote (Figure 2; Dyall et al. 2004). Primary endosymbiosis then gave rise to three lineages; the glaucophytes, red algae, and green algae – the ancestors of land plants (Dyall et al. 2004). Secondary endosymbiosis then gave rise to several other lineages containing secondary plastids surrounded by more than two membranes (Figure 2). Although chloroplasts most closely resemble the the cyanobacterial ancestor to plastids, other plastid types that lack photosynthetic pigments exist in large numbers in land plants. It has been suggested that controlled proliferation of non-green plastids at particular stages in a plant's life cycle and in different organs or tissues is an adaptation to land colonization by plants (Thomson and Whatley 1980). For example, in tissues where photosynthesis would not be possible or non-essential, energy expenditure would be reduced by maintaining plastids in a non-photosynthesizing state while maintaining basic metabolic processes (Thomson and Whatley 1980). Furthermore, the existence of chromoplasts, which contain brightly coloured pigments other than chlorophyll, would have provided angiosperms with the evolutionary advantage of attracting insect or bird pollinators (Thomson and Whatley 1980).

The chloroplast genome of *Arabidopsis thaliana* contains approximately 120 genes, while known cyanobacterial genomes code for at least 1,500 proteins (Raven and Allen 2003). Over evolution, many plastid genes were either lost, or transferred to the nuclear genome of the host. In higher plants, the chloroplast genome encodes mostly proteins involved in translation and photosynthesis; and the approximately 95% of remaining chloroplast proteins are encoded in the nucleus (Martin and Herrmann 1998). In Arabidopsis, genes of cyanobacterial origin account for approximately 18% of all protein-encoding genes in the nuclear genome (Martin et al. 2002). As a consequence of massive gene transfer, a protein trafficking system evolved for delivering

Figure 2. Plastid evolution. Primary endosymbiosis gave rise to three lineages: Glaucophytes, red algae and green algae from which land plants evolved. Secondary endosymbiosis then gave rise to several lineages with secondary plastids possessing more than two membranes. Figure adapted from Keeling, 2004.



Glaucophytes
Red Algae
Green Algae
Land Plants
Cryptomonads
Haptophytes
Heterokonts
Apicomplexa
Ciliates
Euglenids
Chlorarachniophytes
Dinoflagellates
Dinophysis
Karenia
Kryptoperidinium
Lepidodinium

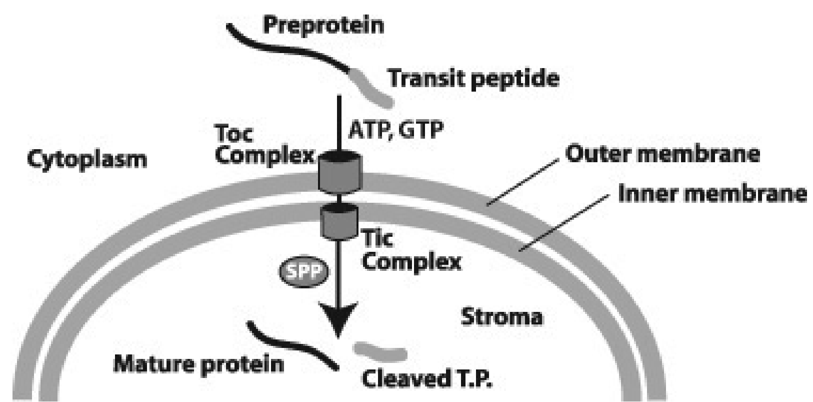
proteins translated in the cytosol back to plastids where they carry out their functions. Gram negative cyanobacteria lack machinery for importing polypeptides, indicating that acquisition of a protein import system evolved following endosymbiosis, coinciding with gene transfer to the nucleus (Reumann, Inoue, and Keegstra 2005). Several components of the import apparatus share homology with cyanobacterial proteins; however, in land plants they have adopted a new function in plastid protein import (Reumann, Inoue, and Keegstra 2005). There are also several protein import components that are eukaryotic in origin, and are thought to have evolved from pre-existing genes of the eukaryotic host (Reumann, Inoue, and Keegstra 2005).

1.3 Chloroplast protein import

Plastid protein import plays a crucial role in chloroplast biogenesis and interconversion between plastid types, as specific plastid types require distinct sets of proteins involved in their unique biochemical pathways. Most nuclear-encoded plastid proteins are translated in the cytoplasm as precursors with a cleavable, N-terminal transit peptide and are imported through the **t**ranslocon at the **o**uter and **i**nner membranes of **c**hloroplasts (Toc and Tic, respectively) (Smith 2006). Receptors of the Toc complex recognize precursors and facilitate the initial steps of import of the unfolded preprotein. In coordination with the Toc complex, the Tic complex completes translocation through the intermembrane space and across the inner envelope membrane into the stroma. Upon emergence in the stroma, the transit peptide is cleaved by a stromal processing peptidase, generating the mature protein (Figure 3; Smith 2006). Some proteins are then further targeted to plastid subcompartments such as the intermembrane space, inner envelope membrane or the thylakoid lumen and membrane (Smith 2006). Multiple targeting pathways exist for targeting to thylakoids and are conserved from the cyanobacterial ancestor (Reumann, Inoue, and Keegstra 2005).

While most proteins that are targeted to the stroma follow the Toc/Tic pathway, several proteins have been identified that are imported without a transit peptide. Two such examples are Tic32, an inner

Figure 3. Toc/Tic protein import pathway. Most nuclear-encoded chloroplast proteins are translated with an N-terminal transit peptide, which facilitates their import into the stroma via the Toc/Tic translocons. Translocation requires energy in the form of GTP and ATP. The transit peptide is subsequently cleaved in the stroma by a stromal processing peptidase (SPP). Figure adapted from Smith and Schnell (2004).


Preprotein
Transit peptide
Toc
Complex
ATP, GTP
Cytoplasm
Outer membrane
Inner membrane
Tic
Complex
SPP
Stroma
Mature protein
Cleaved T.P.

membrane protein thought to be involved in regulation of the Toc/Tic pathway (Nada and Soll 2004), and ceQORH (chloroplast envelope quinone oxidoreductase), which is a peripheral membrane protein bound to the stromal side of the inner membrane (Miras et al. 2007). Both of these proteins are initially imported into the stroma but lack a cleavable transit peptide. While evidence suggests that these proteins do not use the Toc/Tic pathway, the translocon(s) involved in their import have not been identified (Nada and Soll 2004; Miras et al. 2007). An ER to chloroplast protein sorting pathway has also been identified. Proteins that follow this pathway are first targeted to the ER by the presence of a signal peptide and are then targeted to the chloroplast *via* the secretory pathway (Villarejo et al. 2005; Nanjo et al. 2006).

Less common pathways exist for targeting of outer membrane, inner membrane, and intermembrane space-resident proteins. Approximately 24 outer membrane proteins have been identified; most of which lack a transit peptide (Inaba and Schnell 2008). Those that possess alpha-helical transmembrane regions have their targeting information adjacent to and within these regions (Inaba and Schnell 2008). Recently, a cytoplasmic ankyrin repeat protein, AKR2A, has been identified in Arabidopsis that mediates targeting of the outer envelope protein OEP7/14. Evidence suggests that AKR2A may also be involved in targeting of other outer membrane proteins (Bae et al. 2008). Some evidence also exists that Toc75 – the outer membrane channel through which transit peptide-containing precursors are translocated – may also be involved in targeting of outer membrane proteins containing alpha helical transmembrane segments (Inaba and Schnell 2008).

Less information exists on protein targeting to the intermembrane space. So far, targeting of only two intermembrane space proteins have been studied in detail: MGDG synthase and Tic22, both of which interact with the Toc complex but appear to reach the intermembrane space *via* different mechanisms (Inaba and Schnell 2008). There also appears to be at least two distinct targeting mechanisms of inner membrane proteins, both involving transit peptide-mediated interactions with the Toc and Tic complexes. In one pathway, proteins are inserted into the inner membrane by a stop-

transfer signal recognized by the Tic complex. Alternatively inner membrane insertion can occur following complete translocation into the stroma, and requires distinct proteinaceous factors at the inner envelope (Inaba and Schnell 2008).

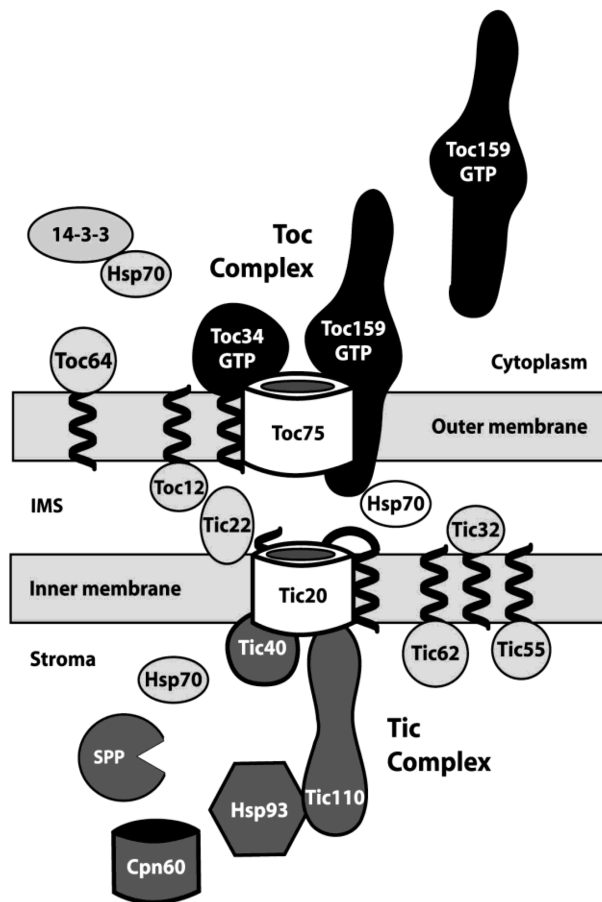
Despite the growing evidence of novel protein targeting pathways to plastids, the majority of chloroplast proteins follow the Toc/Tic pathway. While much work has been done in characterizing this pathway, there are still aspects that remain unclear such as the biogenesis of the Toc/Tic complexes, transit peptide recognition and the molecular events of translocation.

1.4 The general import apparatus

1.4.1 The Tic complex

Figure 4 shows a schematic representation of known components of the Toc and Tic complexes as identified in pea (*Pisum sativum*); the name of each component represents its molecular weight in kilodaltons (Smith 2006). The four components of the Tic complex shown to be directly involved in translocation are Tic20/21, Tic22, Tic110 and Tic40. Tic22 resides in the intermembrane space, and has been suggested to be involved in the formation of Toc/Tic supercomplexes (Kouranov et al. 1998). The inner membrane protein translocation channel is mainly formed by Tic20 and Tic21 (Inaba and Schnell 2008). Both Tic20 and Tic21 are distantly related to the mitochondrial inner membrane channels Tim17 and 23. They share little sequence similarity; however, they are both hydrophobic inner membrane proteins containing four alpha-helical transmembrane segments, and have a similar topology in the chloroplast inner membrane (Inaba and Schnell 2008). Tic20 and Tic21 are differentially expressed throughout the plant, suggesting that they function at different times in development (Teng et al. 2006). It has been proposed that Tic110 also forms part of the inner membrane channel, based on its ability to form a β-barrel channel following reconstitution into proteoliposomes (Heins et al. 2002). However, the portion of the protein involved in the formation of the β-barrel in proteoliposomes lies in the stromal domain of Tic110, suggesting that the protein does not function as a channel *in vivo* (Inaba et al. 2003).

Figure 4. Schematic diagram of the Toc and Tic complexes. Components in black are involved in preprotein recognition, membrane translocation components are shown in white, and components in dark gray are involved in preprotein translocation and maturation of the preprotein (ie transit peptide cleavage and protein folding) in the stroma. Components in light gray have unknown functions or are involved in regulating import under specialized conditions. IMS; intermembrane space. Figure taken from Smith, 2006.


Toc159
GTP
14-3-3
Hsp70
Toc
Complex
Toc34
GTP
Toc159
GTP
Toc64
Cytoplasm
Toc75
Outer membrane
Toc12
IMS
Tic22
Hsp70
Tic32
Inner membrane
Tic20
Tic40
Tic62
Tic55
Stroma
Hsp70
SPP
Tic
Complex
Hsp93
Tic110
Cpn60

Tic110 is encoded by a single gene in most species studied and is the most abundant protein of the Tic complex (Reumann, Inoue, and Keegstra 2005). The Arabidopsis Tic110 knockout mutant is embryo lethal, indicating that Tic110 is essential for plastid biogenesis and import (Inaba et al. 2005). Tic110 contains two short alpha-helical transmembrane segments at its N-terminus and a large soluble domain that protrudes into the stroma (Inaba et al. 2003). The soluble domain binds transit peptides in close proximity to the stromal side of the inner membrane (Inaba et al. 2003). Tic110 also plays a role in the recruitment of the stromal chaperone Hsp93, which also interacts with preproteins during translocation (Nielsen et al. 1997; Kovacheva et al. 2005). ATP hydrolysis by Hsp93 is thought to drive preprotein translocation across the inner membrane and accounts for the stromal ATP requirement of protein import (Kovacheva et al. 2005). Interacting with Tic110 and Hsp93 is the co-chaperone Tic40 (Chou et al. 2003). Tic40 is embedded into the inner membrane by a single transmembrane segment. In addition, it has tetratricopeptide repeat (TPR) and Hip/Hop (Hsc70-interacting protein/Hsc70/Hsp90-organizing protein) domains characteristic of co-chaperones, and is able to stimulate Hsp93 ATP hydrolysis activity (Bedard et al. 2007). Tic110, Tic40 and Hsp93 are found in close association with each other in the Tic complex, and evidence suggests that an association between Tic110 and Tic40 results in binding of Hsp93 to the precursor protein. Rounds of ATP hydrolysis by Hsp93 then facilitate translocation of the preprotein into the stroma (Inaba et al. 2005; Chou et al. 2003; Chou et al. 2006).

Tic62, Tic55 and Tic32 are redox proteins that were found to be associated with the Tic complex (Caliebe et al. 1997; Hormann et al. 2004; Stengel et al. 2008). It has been shown that redox signals regulate the import of ferredoxin-NAD(P)$^+$ oxidoreductase (Yan et al. 2006) and ferredoxin-III (Hirohashi, Hase and Nakai 2001), suggesting that Tic62, Tic55 and Tic32 may regulate this process (Inaba & Schnell 2008). Tic32 is a calmodulin-binding protein and has been shown to be essential for plant viability (Hormann et al. 2004), suggesting a role for calcium in the regulation of protein import.

1.4.2 The Toc complex

Toc75 forms the channel in the outer membrane through which unfolded precursors are translocated (Figure 4; Schnell, Kessler, and Blobel 1994). Evidence in support of this function is the observation that heterologously expressed pea Toc75 forms a β-barrel structure when reconstituted into liposomes with a pore size of approximately 14 to 16 Å – large enough to allow the passage of unfolded preproteins (Hinnah et al. 2002). A homologue of Toc75 (SynToc75) has been identified in the gram-negative cyanobacteria *Synechocystis* sp. PCC 6803 that is structurally similar to Toc75 (Reumann, Inoue, and Keegstra 2005). Both Toc75 and SynToc75 share distant homology with Omp85 proteins, a family of outer membrane proteins found in gram-negative bacteria responsible for protein assembly in the outer membrane (Gentle et al. 2005). Furthermore, Omp85 proteins show homology with channels that secrete virulence factors such as hemolysins and adhesins (Reumann, Inoue, and Keegstra 2005). Although it appears as though gram negative bacteria do not possess translocation machinery for the import of peptides, the import channel most likely evolved following transfer of a cyanobacterial gene encoding an outer membrane protein-secreting channel to the nucleus, and a subsequent reversal of topology in the outer chloroplast membrane (Reumann, Inoue, and Keegstra 2005). Toc75 also exists in a pool not associated with other Toc components (Kouranov et al. 1998).

Toc34 and Toc159 are homologous GTPases that are involved in preprotein recognition at the outer envelope membrane. Toc34 is an integral membrane protein anchored in the chloroplast outer envelope by a single transmembrane segment at the C-terminal end, with a small portion of the C-terminus extending into the intermembrane space (Kessler et al. 1994). The remainder of the protein, which includes the GTP-binding domain, protrudes into the cytosol (Kessler et al. 1994; Tsai, Tu, and Li 1999). Toc159 has a tripartite structure; in addition to the central GTP-binding domain, this receptor also possesses an N-terminal acidic (A-) and a C-terminal membrane anchor (M-) domain (Bauer et al. 2000). Toc159 is anchored into the membrane as shown by protease treatment of isolated chloroplasts (Hirsch et al. 1994); however, the ~52 kDa protease-protected M-domain contains no predicted

transmembrane helices, and the nature of membrane-insertion is unknown (Kessler et al. 1994). It has been speculated that the M-domain may be inserted into the protein channel, Toc75 (Ma et al. 1996). In Arabidopsis, each of Toc75, Toc34 and Toc159 are represented by multigene families (Jarvis et al. 1998; Bauer et al. 2000; Jackson-Constan and Keegstra 2001). Evidence suggests that the Toc34 and Toc159 homologues may also demonstrate functional specificity (see section 1.5).

Two other putative Toc complex components have been identified, Toc64 and Toc12. Toc64 was found to co-fractionate with Toc complex components following solubilisation of outer envelope membranes and cross-linked to other Toc complex subunits (Sohrt and Soll 2000). However, in a separate study, further purification of fractionated Toc complexes led to removal of Toc64 (Schleiff et al. 2003), suggesting that its association with the Toc complex may be transient. In *Physcomitrella patens*, it has been shown that Toc64 is not required for import, and it was suggested to be involved in import under specialized conditions (Hofmann and Theg 2005). Toc12 interacts with Toc64 and Tic22, and has been shown to recruit Hsp70 in the intermembrane space (Becker et al. 2004a), and it has been suggested that this protein might be involved in the formation of Toc-Tic supercomplexes (Smith 2006).

1.5 Multiple protein import pathways

In *Arabidopsis thaliana*, multiple homologues of Toc34 and Toc159 exist, each homologue encoded by different genes (Jarvis et al. 1998; Bauer et al. 2000). There are two homologues of Toc34 in Arabidopsis – atToc33 and atToc34 (the letters in front of each component designate their species of origin). Toc159 has four homologues in Arabidopsis, namely atToc159, atToc132, atToc120 and atToc90. Characterization of Arabidopsis Toc component knockout mutants indicates that homologues within both the Toc34 and Toc159 families may demonstrate functional specificity. The Toc159 homologues have different expression profiles; atToc159 is present at approximately 10-fold higher levels than atToc132/120 in green tissues, however, all homologues show relatively low uniform levels

of expression in non-green tissues (Kubis et al. 2004). The atToc159 knockout mutant, *ppi*2 (plastid protein import mutant 2) has a severe albino phenotype and chloroplast biogenesis in this mutant is blocked early in development (Bauer et al. 2000). The *ppi*2 mutant can be partially rescued with sucrose supplementation, and photosynthetic genes are specifically down-regulated in *ppi*2 (Bauer et al. 2000), whereas expression of several chloroplast proteins not involved in photosynthesis was normal in *ppi*2 (Bauer et al. 2000). Furthermore, when transiently expressed in *ppi*2, a photosynthetic transit peptide-GFP fusion protein (Rubisco small subunit transit peptide fused to GFP) showed decreased import relative to a non-photosynthetic transit peptide-GFP fusion protein (pyruvate dehydrogenase E1α subunit transit peptide fused to GFP). Together, these data suggest that atToc159 is involved in the import of photosynthetic proteins. Consistent with this hypothesis, atToc159 was found to specifically bind the transit peptide of photosynthetic preproteins during *in vitro* solid-phase binding assays (Smith et al. 2004).

AtToc132 and atToc120 share more sequence identity with each other than with atToc159, and it has been suggested that these receptors are functionally redundant (Bauer et al. 2000; Ivanova et al. 2004). Knockout mutants of either atToc132 or atToc120 have less severe phenotypes compared to *ppi2* (Ivanova et al. 2004; Kubis et al. 2004). However, double knockout mutants of atToc132 and atToc120 are non-viable, and overexpression of atToc132 fails to rescue the *ppi*2 mutant (Ivanova et al. 2004; Kubis et al. 2004), indicating that these receptors have a function distinct from that of atToc159. *In vitro*, atToc132 was shown to selectively bind the transit peptide of a representative non-photosynthetic protein (Ivanova et al. 2004). *AtTOC90* has a similar expression profile to *atTOC159* and also appears to be involved in the import of photosynthetic proteins (Hiltbrunner et al. 2004).

The Toc34 homologues in Arabidopsis also show different developmental expression profiles. *atTOC33* is expressed at very high levels in young, rapidly expanding photosynthetic tissues, whereas *atTOC34* is expressed at low levels throughout development (Jarvis et al. 1998; Gutensohn et al. 2000; Kubis et al. 2003). Analysis of the atToc33 and atToc34 knockout mutants has provided evidence that

these receptors also contribute to the functional specificity of the Toc complex. The atToc33 knockout mutant (*ppi*1) has a pale phenotype during the first two weeks of development, suggesting a role in the import of photosynthetic proteins. Similar to the *ppi*2 mutant, *ppi*1 also shows specific downregulation of photosynthetic genes (Kubis et al. 2003). Furthermore, isolated *ppi*1 chloroplasts were specifically defective in the import of representative photosynthetic proteins in comparison to a non-photosynthetic protein (Kubis et al. 2003). The atToc34 knockout mutant (*ppi*3) shows normal development of green tissues but root growth is slightly retarded (Constan et al. 2004). Despite the apparent effect of this mutation on root development, root plastids appeared very similar to those of wild-type Arabidopsis (Constan et al. 2004). The chloroplasts of *ppi*3 are able to import photosynthetic proteins with a similar efficiency to wild-type chloroplasts (Constan et al. 2004); however, the efficiency of nonphotosynthetic protein import has not been investigated. The Toc34 receptors in Arabidopsis do show functional overlap as evidenced by the viability of their respective knockout mutants (Jarvis et al. 1998; Constan et al. 2004). Furthermore, overexpression of atToc34 under the control of a constitutive promoter is able to fully complement the *ppi*1 mutant (Jarvis et al. 1998).

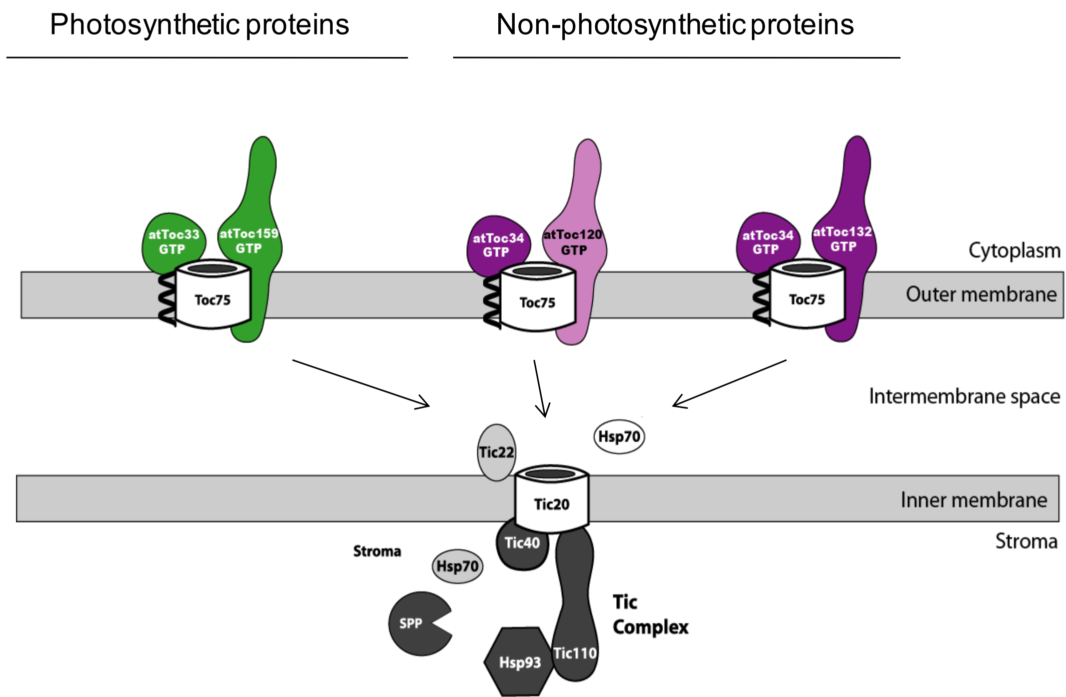
The functional specificity between members of the Toc34 and Toc159 families in Arabidopsis strongly suggests that multiple protein import pathways exist that are able to differentially import photosynthetic versus non-photosynthetic chloroplast proteins. Co-immunoprecipitation studies with detergent extracts of total chloroplast membranes demonstrated that atToc159 forms Toc complexes distinct from atToc132 or atToc120 complexes (Ivanova et al. 2004). Furthermore, atToc132 and atToc120 were found to co-immunoprecipitate, providing further evidence that atToc132 and atToc120 represent a functionally redundant subclass of recepetors (Ivanova et al. 2004). It was also demonstrated that atToc33 mainly co-immunoprecipitates with atToc159, while atToc34 co-immunoprecipitates mainly with atToc132/120 (Ivanova et al. 2004). This led to the hypothesis that structurally distinct Toc complexes exist consisting of atToc159/33 or atToc132/120/34 (Ivanova et al. 2004). The Toc34 homologues did show some cross-reactivity with atToc159- and atToc132/120- containing complexes

in these co-immunoprecipitation studies (Ivanova et al. 2004), which is consistent with the observed functional redundancy between atToc33 and atToc34 (Jarvis et al. 1998). Arabidopsis has only one functional Toc75 homologue that co-immunoprecipitates with both atToc159/33- and atToc132/120/34-containing complexes (Ivanova et al. 2004). Furthermore, these functionally distinct import pathways seem to converge at the Tic complex, as Tic110 appears to be involved in both pathways (Figure 5). However, Tic20 and Tic21 show differential expression which may suggest that they function in distinct import pathways (Inaba and Schnell 2008).

So far, the existence of structurally and functionally distinct Toc complexes has been shown only in Arabidopsis. However, evidence is accumulating that multiple import pathways may exist in other species as well. Using bioinformatic approaches, multiple Toc159 homologues have also been found in rice (*Oryza sativa*; Kubis et al. 2004) and Toc34 homologues with distinct expression patterns have been identified in spinach (Voigt et al. 2005) and spruce (*Picea abies* L. Karst; Fulgosi et al. 2005). In addition, Voigt et al. (2005) report that multiple homologues of Toc34 have been identified by searching available expressed sequence tags in rape seed, potato, tomato and maize.

The existence of structurally and functionally distinct Toc complexes is thought to be a mechanism for plastids to adapt to changing gene-expression profiles during plant and plastid development. Multiple import pathways allow plastids to quickly and efficiently balance import of plastid house-keeping proteins and highly expressed proteins required for specialized metabolic functions (i.e. photosynthetic proteins) (Inaba and Schnell 2008). The dyanamic nature of plastids and their ability to interconvert between types, require mechanisms of adaptability such as those provided by multiple protein import pathways. The discovery of structurally and functionally distinct import complexes raises several questions including 1) how are structurally distinct Toc complexes formed, and 2) how are these structurally distinct Toc complexes able to discriminate between different classes of preproteins.

Figure 5. Multiple protein import pathways in Arabidopsis. Structurally distinct Toc complexes exist in Arabidopsis that demonstrate functional specificity. Toc complexes containing atToc159 and atToc33 are involved in the import of photosynthetic proteins, whereas atToc132/120- and atToc34- containing complexes import mainly non-photosynthetic, constitutively expressed plastid proteins. atToc75 is common to both pathways, and these pathways appear to converge at the Tic complex.


Photosynthetic proteins
Non-photosynthetic proteins
atToc33
GTP
atToc159
GTP
atToc34
GTP
atToc120
GTP
atToc34
GTP
atToc132
GTP
Cytoplasm
Toc75
Toc75
Toc75
Outer membrane
Intermembrane space
Hsp70
Tic22
Tic20
Inner membrane
Stroma
Tic40
Stroma
Hsp70
SPP
Tic
Complex
Hsp93
Tic110

1.6 The Arabidopsis Toc159 family of receptors

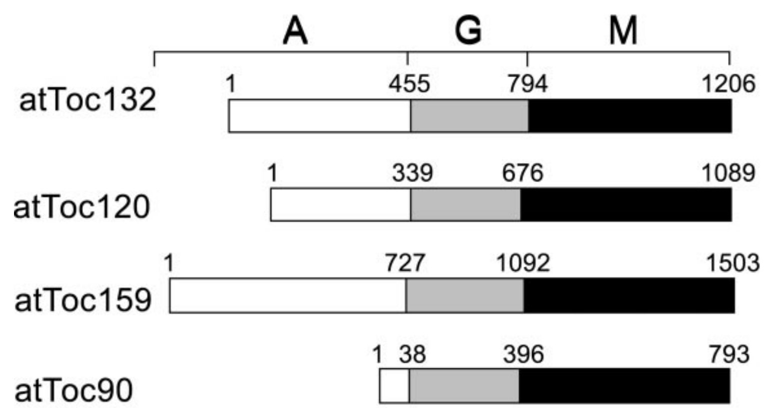
The Arabidopsis Toc159 homologues have a tripartite structure; in addition to a central GTPase (G-) domain, they have a C-terminal membrane anchor (M-), and an N-terminal acidic (A-) domain (Figure 6; Bauer et al. 2000). The G- and M-domains share approximately 65% sequence identity between receptors. The A-domains are more variable in length and sequence (~20% sequence identity; Ivanova et al. 2004). Due to its variability between homologues, the A-domain has been proposed to contribute to the functional specificity of these receptors. Removal of the A- and G-domain of the Toc159 receptors by protease shaving of cytosol-exposed proteins in isolated chloroplasts does not significantly impair import *in vitro* (Chen, Chen, and Schnell 2000). Furthermore, overexpression of the M-domain of atToc159 alone can partially restore the import of photosynthetic proteins in the atToc159 knockout mutant (Lee et al. 2003). Together, these findings suggest that the A- and G-domains, while not essential for import, may serve a regulatory function. The A-domain contains no conserved functional domains and its function remains elusive.

1.6.1 Targeting of Toc159 to chloroplasts during initial Toc complex assembly

Components of the Toc complex are encoded in the nucleus and during initial Toc complex assembly are targeted to the chloroplast outer envelope membrane. Of the core components of the Toc complex, targeting of Toc75 is the most well-characterized. Toc75 is the only known outer membrane protein of chloroplasts that is synthesized as a higher molecular weight precursor, and undergoes multiple cleavages during targeting (Tranel et al. 1995; Schleiff and Klösgen 2001). The transit peptide of Toc75 is bipartite – the most N-terminal portion mediates translocation of Toc75 into the stroma (*via* the Toc/Tic pathway, through Toc75) and is removed by a stromal processing peptidase (Tranel and Keegstra 1996). The C-terminal portion of the transit peptide includes a polyglycine stretch that acts as a stop-transfer signal and is cleaved by a type I signal peptidase in the outer envelope membrane (Inoue et al. 2005).

Figure 6. Amino acid sequence comparison of the Arabidopsis Toc159 receptors. There are four Toc159 homologues in Arabidopsis, each possessing a tripartite structure. (A) is a schematic representation of the Arabidopsis Toc159 homologues delineating the acidic (A), GTP-binding (G), and membrane anchor (M) domains. The number above each receptor represents the bordering amino acid residue between each domain. The sequence identities of each domain between each receptor and atToc132 are shown in (B).

A



B

atToc132	A	G	M
atToc120	36	93	87
atToc159	11	49	49
atToc90	-	41	32

The transmembrane domain and C-terminal tail are required for proper targeting of the Toc34 homologues. These regions are able to target a carrier protein to the chloroplast outer membrane (Chen and Schnell 1997). The energy requirements of Toc34 targeting are controversial (Kessler et al 1994; Seedorf, et al. 1995; Chen and Schnell 1997; Tsai, Tu, and Li 1999; Qbadou, et al. 2003); however, it appears as though GTP/GDP plays a role in its insertion into the outer membrane. In addition, it has been suggested that the type of lipid (Qbadou, et al. 2003) and interactions with other proteinaceous factors (Tsai, Tu, and Li 1999) are also involved in targeting of Toc34. AKR2A, a recently identified protein possessing chaperone activity has been shown to be involved in targeting of the chloroplast outer envelope protein, OEP7, to the outer membrane (Bae et al. 2008). AKR2A also binds atToc33 and atToc34; however it has not yet been shown to play a direct role in targeting of the Toc34 proteins to chloroplasts (Bae et al. 2008).

In vitro targeting assays using isolated chloroplasts have shown that atToc159 binding to chloroplasts does not require energy (Smith et al. 2002). However, insertion of atToc159 is stimulated by GTP, and the receptor is more easily inserted in its GDP-bound form (Smith et al. 2002b). Furthermore, disruption of the GTP hydrolysis activity or GDP-binding capacity of this receptor decreases insertion (Smith et al. 2002b; Bauer et al. 2002). In addition to the GTPase activity of atToc159, insertion is also stimulated by the GTPase activity of atToc33 (Wallas et al. 2003), and *in vitro* protein-protein interaction studies have demonstrated a specific interaction between the G-domains of atToc159 and atToc33 (Hiltbrunner et al. 2001; Smith et al. 2002b; Bauer et al. 2002; Ivanova et al. 2004). The nature of insertion of the Toc159 receptors into the outer membrane is unclear; however, it has been shown that insertion is critical for the formation of a functional Toc complex able to import preproteins (Chen, Chen, and Schnell 2000; Lee et al. 2003). The M-domain possesses no predicted transmembrane helices, yet has been shown to be embedded in the outer membrane by its protection from protease treatment and alkali-extraction (Kessler et al. 1994; Bauer et al. 2000). The M-domain of Toc159 is involved in the translocation event and it has been proposed to

contribute to the structure and function of the translocation channel (Ma et al. 1996). In support of this hypothesis, integration of atToc159 into proteoliposomes requires both atToc33 and atToc75 (Wallas et al. 2003). Furthermore, while the M-domain binds to isolated chloroplasts on its own, the G-domain is required for its insertion (Wallas et al. 2003).

1.6.2 Preprotein recognition by the Toc159 receptors

Based on its ability to cross-link with the transit peptide of precursor proteins during import into pea chloroplasts, it was proposed that Toc159 acts as the primary preprotein receptor of the Toc complex (Perry and Keegstra 1994; Ma et al. 1996; Kouranov and Schnell 1997). Following the observation that Toc159 exists as a multigene family in Arabidopsis, characterization of the atToc159 knockout mutant suggested that this receptor is involved in the import of photosynthetic proteins, whereas atToc132/120 may import nonphotosynthetic proteins, or proteins required for basic plastid functioning (Bauer et al. 2000; Ivanova et al. 2004; Kubis et al. 2004). It has been shown using *in vitro* solid-phase binding studies that the Toc159 receptors are able to bind transit peptides, with different specificities; atToc159 with a photosynthetic transit peptide fusion protein (Smith et al. 2004), and atToc132/120 with a nonphotosynthetic transit peptide fusion protein (Ivanova et al. 2004).

It is unclear how transit peptides facilitate recognition of preproteins by receptors of the Toc complex. In general, transit peptides are rich in hydroxylated and hydrophobic amino acid residues, and lack acidic residues (von Heijne, Steppuhn, and Herrmann 1989). In addition, they are variable in length and their primary structures show little sequence conservation (von Heijne, Steppuhn, and Herrmann 1989). They have no known secondary or tertiary structures that contribute to recognition by receptors of the Toc complex (Bruce 2001). In aqueous solvents, transit peptides are largely unstructured; however, alpha-helical structure can be induced in a chloroplast outer membrane-mimicking environment (Bruce 2001). It has been suggested that a conformational change in transit peptides while in close proximity to the outer envelope membrane may facilitate its recognition by the

Toc complex receptors; however, this has not been shown experimentally. This mechanism parallels the mitochondrial presequence, whose alpha-helical structure is essential for its recognition by receptors of the **t**ranslocon at the **o**uter mitochondrial **m**embrane (Tom complex; Bruce 2001). Several attempts have been made to identify specific sequences within the transit peptide of pSSU (Rubisco small subunit precursor) required for import (von Heijne, Steppuhn, and Herrmann 1989; Lee et al. 2002; Lee et al. 2006). Despite these efforts, it appears that the overall context of the transit peptide is important for proper targeting, and also that there may be regions of overlapping function within the transit peptide of pSSU. A cross-linking approach was taken to try to identify regions of atToc159 involved in transit peptide binding, demonstrating that the transit peptide of pSSU cross-linked to the G- and M-domains (Smith et al. 2004). While a role for the A-domain in preprotein binding has not previously been observed, the variability within this domain may provide an explanation for the ability of the Toc159 receptors to recognize such a large number of structurally diverse transit peptides with varying specificity.

1.7 Overall objectives

The Toc159 receptors have been shown to be the major chloroplast preprotein receptors and directly contribute to the functional specificity of structurally distinct Toc complexes involved in multiple chloroplast protein import pathways. Previous studies have mainly focused on determining the function of the G-domain of this receptor. The main goal of the current study was to shed light onto the function(s) of the A-domain of the Arabidopsis Toc159 receptors in an attempt to further elucidate the structural determinants of Toc159 receptor specificity. This thesis is presented in two chapters; each addressing the following specific objectives of the current research: 1) to investigate the role of the A-domains in targeting of the Toc159 receptors to structurally distinct Toc complexes during initial Toc complex assembly, and 2) to perform a structural analysis of the Toc159 family A-domains which will

aid in making inferences about its function. The findings of each chapter will be discussed in an overall conclusion of this study.

2. Investigation of the role of the A-domain in targeting the Toc159 receptors to structurally distinct Toc complexes

2.1 Background

The observation that distinct Toc complexes exist in Arabidopsis raises the question of how they are initially formed. The Toc159 family members are most similar over their G- and M-domains, which suggests that the G-domain is involved in targeting of each of the Toc159 receptors in a mechanism similar to that used by atToc159. This is supported by the observation that atToc132 and atToc120 also bind the G-domain of Toc34 homologues *in vitro* (Ivanova et al. 2004). The structural determinants of the specificity of Toc159 targeting are unknown. It has been proposed that the A-domain contributes to the functional specificity due to its high level of variability in primary sequence among receptors of this family (Bauer et al. 2000). Previous targeting studies have shown that the A-domain of atToc159 is not required for targeting to isolated chloroplasts, as its removal does not affect binding or insertion of the receptor into the outer membrane. Furthermore, atToc90, which lacks an A-domain is able to target to isolated chloroplasts (Smith et al. 2002b; Hiltbrunner et al. 2004). Targeting of other Toc159 homologues has not been characterized and it remains unknown how these receptors are able to assemble into structurally and functionally distinct Toc complexes. A role for the A-domain in the specificity of Toc159 targeting has not been investigated.

2.2 Objectives and Hypotheses

The objective of this portion of the study was to investigate the role of the A-domain in targeting of the Toc159 family of receptors to structurally distinct Toc complexes. It was hypothesized, due to the relatively high level of variability found among the A-domains of this family, that this domain would play a role in targeting of the Toc159 receptors to chloroplasts in a way that contributes to the formation of structurally distinct Toc complexes.

2.3 Materials and Methods

2.3.1 Plant material and growth conditions

Plants were grown as outlined in Smith et al. (2002a). Seeds of wild-type (ecotype Columbia), *ppi*1 and *ppi*3 *Arabidopsis thaliana* were surface sterilized by washing in 95% ethanol for 5 min, 30% bleach with 0.02% (v/v) Triton-X 100 for 20 min, and subsequently washed 5 times with sterile water. Approximately 25 mg of seeds per plate were sown on 150 mm x 15 mm plates containing 0.5x Murashige and Skoog media supplemented with 1% sucrose and 0.8% agar. Seeds were stratified at 4°C for 24 h. Plants were grown for 14 to 18 days at 22°C under a 16:8 hour, light:dark cycle in a controlled growth chamber (Enconair, Bigfoot Series).

2.3.2 Isolation of intact chloroplasts from Arabidopsis thaliana

Intact chloroplasts from wild-type, *ppi*1 and *ppi*3 *A. thaliana* were isolated using the following methods adapted from Brock et al. (1993) and Schulz et al. (2004). All buffers were chilled on ice prior to the isolation procedure. Tissue was harvested from 14-18 day *A. thaliana* plants with a typical total fresh weight between 20-30 g. The tissue was homogenized in ice-cold grinding buffer (50 mM HEPES-KOH pH 7.5, 330 mM sorbitol, 2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 0.1% [w/v] ascorbic acid, 0.25% [w/v] BSA) using a PowerGen Homogenizer (Fisher Scientific). The homogenate was filtered through 2 layers of Miracloth (Calbiochem) into a pre-chilled 500 ml centrifuge bottle, and was centrifuged for 8 min, at 4°C, 1,000 g (Beckman Coulter Avanti J-30I centrifuge, JLA 10.5 rotor). The supernatant was decanted and the pellet resuspended in 8 ml fresh, cold grinding buffer. The chloroplast suspension was layered evenly onto two Percoll step gradients consisting of a lower 85% Percoll layer (50 mM HEPES-KOH pH 7.5, 330 mM sorbitol, 2 mM $MgCl_2$, 4 mM EDTA, 0.2% [w/v] BSA, 50 mM ascorbic acid) and an upper 35% Percoll layer (50 mM HEPES-KOH, pH 7.5, 330 mM sorbitol, 2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 50 mM ascorbic acid). Gradients were centrifuged

in a swinging-bucket rotor (JS13.1) at 4°C and 7,700 g for 15 min with slow acceleration and deceleration. The top layer of grinding buffer, broken chloroplasts and a portion of the 35% Percoll layer were aspirated, leaving a layer of intact chloroplasts at the 85%:35% Percoll interface. Intact chloroplasts were transferred to a chilled 50 ml round-bottom centrifuge tube containing approximately 20 ml of cold HS buffer (25 mM HEPES-KOH pH 7.5, 330 mM sorbitol). Cold HS buffer was added to a final volume of 45 to 50 ml. The chloroplast suspension was centrifuged for 6 min at 4°C, 1,000 g (JS 13.1 rotor). The supernatant was decanted and the intact chloroplast pellet was resuspended in approximately 200 to 300 µl of HS buffer.

For targeting assays using wild-type *A. thaliana* chloroplasts, chlorophyll concentration was measured as described previously (Arnon 1949). Briefly, 10 µl of the intact chloroplast suspension were added to 990 µl of 80% acetone. The solution was mixed by vortexing and centrifuged at maximum speed for 2 min to pellet any insoluble material. The absorbance was measured at 652 nm using a Cary 50 Conc UV/Visible Spectrophotometer. The chlorophyll content was determined as follows: chlorophyll concentration (mg/ml) = $[A_{652}/36]$ x DF; where A_{652} is the absorbance measured at 652 nm and DF is the dilution factor of the chloroplast sample (DF=100). Freshly isolated, intact chloroplasts were diluted in HS buffer to a final chlorophyll concentration of 1 mg/ml.

For targeting assays using *ppi*1 and *ppi*3 chloroplasts, protein concentrations of intact chloroplast samples were measured using the Bio-Rad Protein Assay (Bio-Rad), and chloroplasts were diluted to a protein concentration of 9.7 mg/ml, which corresponds to the equivalent of 1 mg/ml chlorophyll in wild-type chloroplasts.

2.3.3 In vitro translation of radiolabelled A. thaliana Toc159 homologues

Constructs used as templates for transcription and translation of radiolabelled *A. thaliana* Toc159 homologues are as follows: pET21d:atToc159 (Smith et al. 2002b), pET21a:atToc132 (Bauer et al. 2000; Ivanova et al. 2004), pET21a:132GM, pET21d:159GM (Smith et al. 2002b),

pET21d:159A132GM, and pET21a:132A159GM (generous gifts from D. Schnell, University of Massachusetts). Radiolabelled receptors were generated using the T_NT Coupled Reticulocyte Lysate System (Promega) according to the manufacturer's directions. Briefly, 1 µg of plasmid DNA encoding a Toc159 construct was used per 50 µl reaction. [^{35}S]Methionine (EXPRES^{35} S^{35} Protein Labeling Mix, PerkinElmer) was added in place of unlabelled methionine to the system, to generate radiolabelled translation products. Other components were added as outlined in the manufacturer's instructions. The reaction mixture was incubated at 30°C for 90 min, and stored at -80°C until further use.

2.3.4 In vitro targeting assays

In vitro targeting assays were performed essentially as described in Smith et al. (2002a). Each targeting reaction included fresh, intact chloroplasts corresponding to the wild-type equivalent of 50 µg of chlorophyll, 1 mM DTT, 1 mM GTP, 2 mM ATP, 10 mM methionine, 50 mM HEPES-KOH pH 7.5, 330 mM sorbitol, 5 mM magnesium acetate, 25 mM potassium acetate, and 4 µl of *in vitro* translated, radiolabelled protein in a final volume of 100 µl. Prior to adding the radiolabelled Toc159 homologue, reaction components were equilibrated by incubation at 26°C for 5 min. After equilibration, radiolabelled protein was added and the targeting reaction was incubated at 26°C for 30 min. The reaction was stopped after 30 min by dilution in 400 µl of ice-cold HS buffer. The reaction mixture was then centrifuged at 2,000 g for 5 min and the chloroplast pellet was gently resuspended in 100 µl of HS buffer. Chloroplasts were then supplemented with 2 mM $CaCl_2$ and treated either with or without 100 µg/ml thermolysin on ice for 30 min. Following incubation, 10 mM EDTA was added to all tubes to stop proteolysis. The chloroplast suspensions were then each layered on 800 µl of 35% Percoll (50 mM HEPES-KOH, pH 7.5, 330 mM sorbitol, 2 mM EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 50 mM ascorbic acid) containing 10 mM EDTA, and reisolated by centrifugation at 6,000 g for 5 min. The supernatant was discarded and the chloroplasts were hypotonically lysed by resuspending in 800 µl of ice cold, 2 mM EDTA, vortexing briefly and incubating on ice for 10 min. NaCl was added to the lysed

chloroplasts to a final concentration of 250 mM. Lysed chloroplasts were centrifuged for 30 min at 17,000 g, 4°C to pellet chloroplast membranes. The membrane pellet was resuspended in 30 µl of 2x SDS-PAGE sample buffer (0.35 M Tris, 2% bromophenol blue, 5% SDS, 0.16 M DTT, 7.5% glycerol). Each targeting reaction was done in duplicate, and each assay was repeated at least 3 times.

2.3.5 SDS-PAGE and phosphorimager analysis

Proteins associated with isolated chloroplast membranes were resolved using SDS-PAGE (4% stacking and 10% resolving gels) under constant current (15 mA through the stacking gel and 25 mA through the resolving gel). Two samples of *in vitro* translation product were also loaded on the gel, each corresponding to $^{1}/_{20}$ of the amount used in each targeting reaction for quantitation purposes. Following electrophoresis, the gel was stained with Coomassie blue (0.25% (w/v) Coomassie Blue R250, 50% (v/v) methanol, 10% (v/v) acetic acid) and dried on a gel dryer (Hoefer Scientific, San Francisco) for 60 min at 75°C. The dried gel was exposed to a phosphor screen (Bio-Rad Laboratories Ltd.) for 2 to 7 days. The exposed phosphor screen was scanned using a phosphorimager (Personal Molecular Imager FX, Bio-Rad Laboratories Ltd.), and levels of radioactivity were quantitated using Quantity One 1-D Analysis software v4.6 (Bio-Rad Laboratories Ltd.). Levels of radioactivity were corrected for the number of methionines present in each radiolabelled protein and binding and insertion efficiencies were calculated as a percentage of the initial amount of radiolabelled receptor added to the targeting reaction. In addition, the corrected radioactivity counts were used to calculate the ratio of inserted:bound receptor. Refer to appendix 1 for *in vitro* targeting assay data.

2.3.6 Statistics

A one-way analysis of variance was used to test for significant differences in binding or insertion between three or more constructs. Student's t-tests (two-sample, assuming unequal variances) were used to make direct comparisons between binding or insertion of two constructs or two groups of

constructs. Significant differences were reported with a P value ≤ 0.05. Statistical analyses are summarized in appendix 2.

2.4 Results

To investigate a general role of the Arabidopsis Toc159 family A-domains in Toc complex assembly, radiolabelled constructs were targeted to chloroplasts isolated from wild-type (WT) Arabidopsis. Radiolabelled constructs used for targeting assays were full-length atToc159 and atToc132, truncated versions of these receptors lacking their respective A-domain (159GM and 132GM) as well as A-domain-swapped constructs (132A159GM and 159A132GM) (Figure 7). As it is the effect of the A-domain on targeting being investigated, those constructs containing the GM-domains of atToc159 may be referred to as "atToc159-derived constructs", while those containing the GM-domains of atToc132 may be referred to as "atToc132-derived constructs". In addition, to determine if the A-domain plays a role in the specificity of Toc159 receptor targeting to structurally distinct Toc complexes, these constructs were also targeted to chloroplasts isolated from the atToc33 knockout (*ppi*1) and atToc34 knockout (*ppi*3) mutants, *in vitro*. The targeting assay method used in this study has been used previously to study atToc159 targeting to isolated chloroplasts (Hiltbrunner et al. 2001; Bauer et al. 2002; Smith et al. 2002b; Wallas et al. 2003).

2.4.1 Effect of the A-domain on targeting to wild-type chloroplasts

In addition to bands that correspond to full-length receptor, targeting assays using atToc159, atToc132, 159A132GM and 132A159GM also produce an ~86 kDa band (Figure 8, lanes 2, 4, 7, 9, 12, 14). This fragment of Toc159 has been previously shown to result from non-specific proteolysis of the A-domain (Bölter, May, and Soll 1998; Chen, Chen, and Schnell 2000). Therefore, both full-length receptor and the ~86 kDa degradation products were included when calculating binding efficiency of atToc132, atToc159, 132A159GM, and 159A132GM to isolated chloroplasts.

Removal of the A-domain of atToc159 (159GM), or addition of the A-domain of atToc132 to 159GM (132A159GM) had no significant effects on binding to WT chloroplasts ($F_{(2,19)} = 0.57$; d.f.= 2, $P = 0.56$) (Figure 9A). This suggests that the A-domain does not play a role in binding of atToc159 to

Figure 7. Schematic representation of recombinant Toc159 receptor constructs used for *in vitro* targeting assays. Constructs were radiolabelled by incorporation of ^{35}S-methionine.

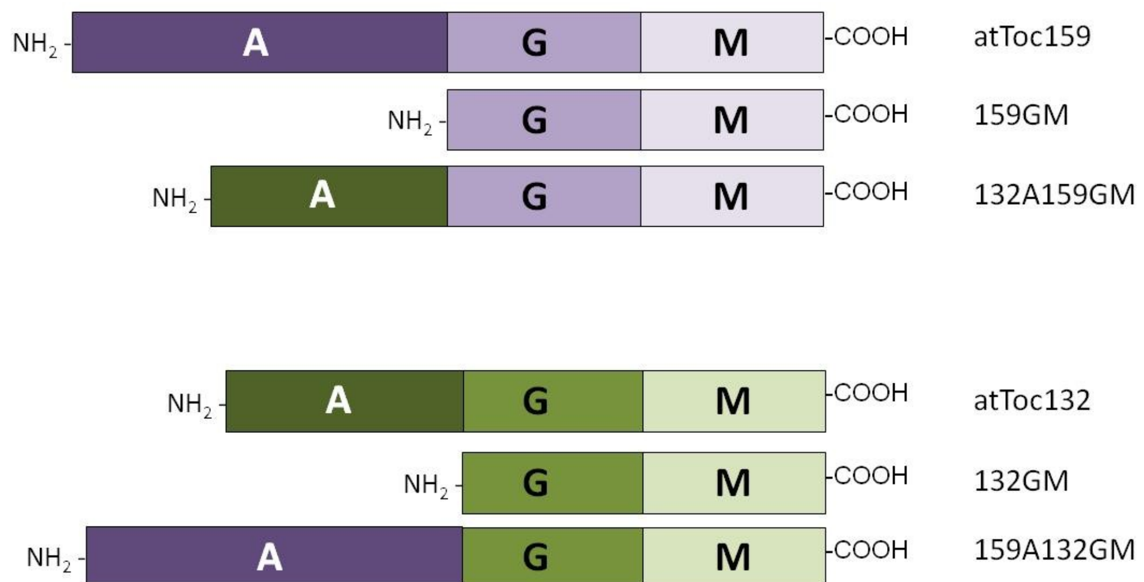
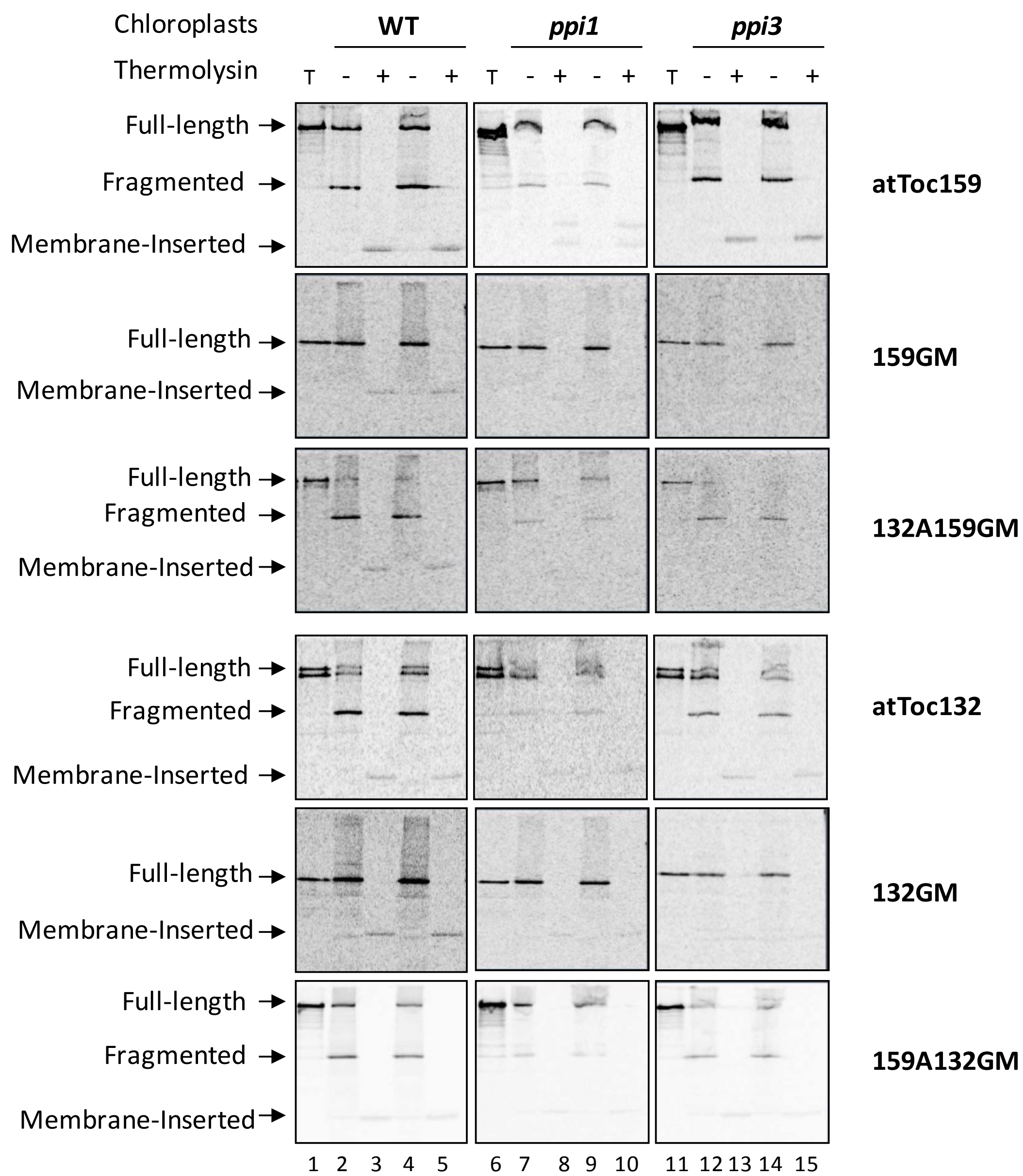

NH2
A
G
M
-COOH
atToc159
NH2
G
M
-COOH
159GM
NH2
A
G
M
-COOH
132A159GM
NH2
A
G
M
-COOH
atToc132
NH2
G
M
-COOH
132GM
NH2
A
G
M
-COOH
159A132GM

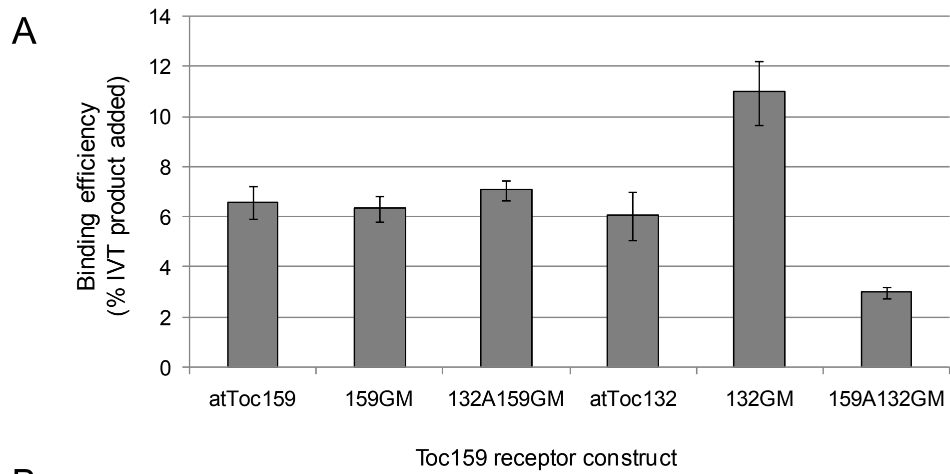
Figure 8. Phosphorimager-visualized SDS-PAGE gels following in vitro targeting assays (see text for details). Lanes 1, 6 and 11 are loaded with *in vitro* translated, radiolabelled receptor corresponding to $^1/_{20}$ of the amount added to each targeting reaction (T). Lanes 2, 4, 7, 9, 12, 14 contain untreated chloroplast membranes with associated bound receptor. Lanes 3, 5, 8, 10, 13, 15 contain chloroplast membranes isolated following thermolysin treatment of intact chloroplasts, showing membrane insertion of the Toc159 receptor constructs.



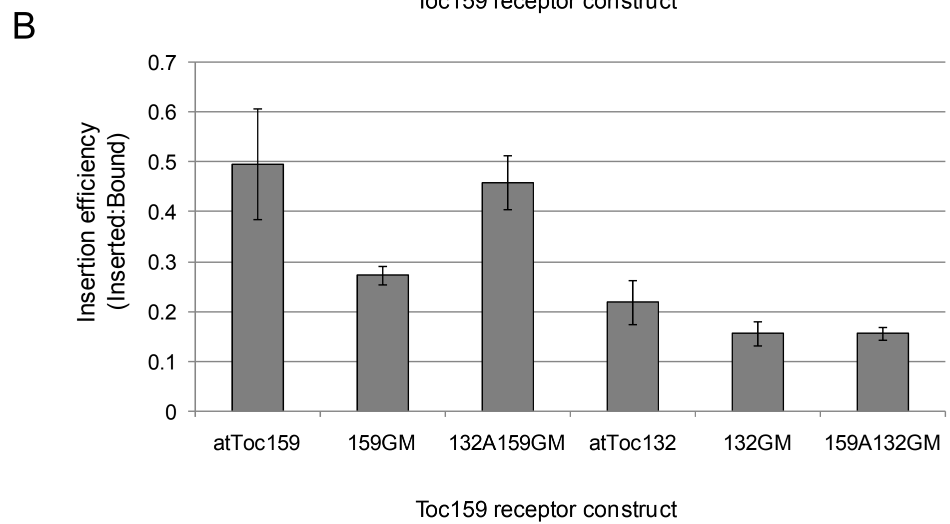
Chloroplasts
WT
ppi1
ppi3
Thermolysin
T - + - + T - + - + T - + - +
Full-length
Fragmented
Membrane-Inserted
atToc159
Full-length
Membrane-Inserted
159GM
Full-length
Fragmented
Membrane-Inserted
132A159GM
Full-length
Fragmented
Membrane-Inserted
atToc132
Full-length
Membrane-Inserted
132GM
Full-length
Fragmented
Membrane-Inserted
159A132GM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Figure 9. Targeting efficiency of Toc159 receptor constructs to chloroplasts isolated from WT plants. Binding efficiency (A), inserted:bound receptor (B) and insertion efficiency (C) to WT chloroplasts were quantitated. Insertion efficiency is represented by the ratio of inserted to bound receptor. Error bars represents standard error of the mean. IVT product; *in vitro* translation product.

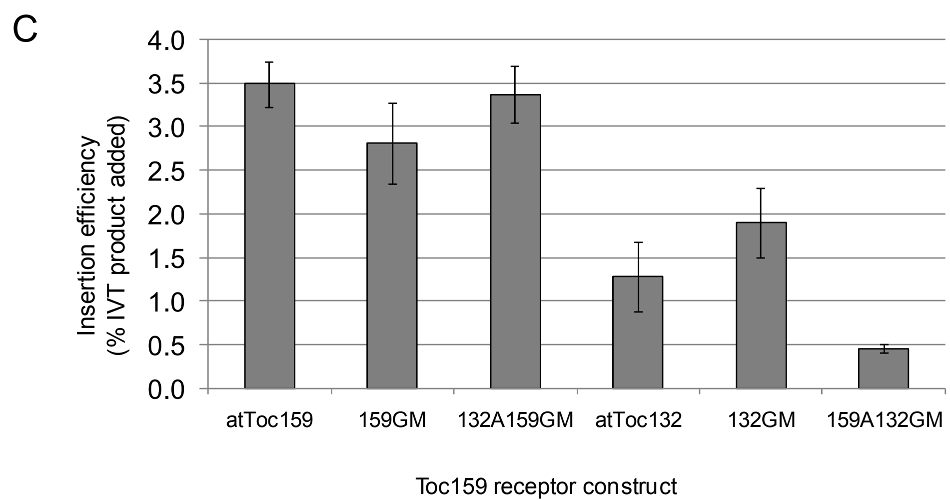
A



B



C



isolated chloroplasts. These results are consistent with previous reports that removal of the A-domain does not have an effect on binding of atToc159 (Smith et al. 2002b). In contrast, significant differences were observed between the atToc132-derived constructs ($F_{(2,18)} = 21.3$, $P < 0.001$). Specifically, removal of the A-domain from atToc132 (132GM) resulted in a significant increase in binding ($t = 2.7$, d.f. = 11, P = 0.018) and addition of the A-domain of atToc159 to atToc132 (159A132GM) significantly decreased the binding efficiency (t = 3.93, d.f. = 11, P = 0.002) relative to atToc132 (Figure 9A). These data point to an inhibitory affect of the A-domain on atToc132 binding to WT chloroplasts, *in vitro*.

Formation of a functional Toc complex capable of importing preproteins requires the insertion of Toc159 into the outer envelope membrane as part of Toc complex assembly (Chen, Chen, and Schnell 2000; Lee et al. 2003). To assess insertion of the constructs, intact chloroplasts were treated with thermolysin following the targeting assay. Thermolysin is a non-specific protease that degrades any proteins found at the periphery of the chloroplast. Treatment with thermolysin gives rise to a characteristic 50-55 kDa protease-protected fragment of the Toc159 homologues, corresponding to the membrane-inserted (M-) domain (Figure 8, lanes 3, 5, 8, 10, 13, 15) (Hirsch et al. 1994; Bauer et al. 2000). In the assay used for this study, the amount of receptor targeted, includes receptor that is bound to the outer membrane as well as receptor that has successfully been inserted into the outer membrane, which are indistinguishable without thermolysin treatment. In this study, insertion was reported as both the percentage of *in vitro* translation product added to the targeting reaction, as well as the ratio of inserted to bound receptor, as determined by thermolysin treatment.

The atToc132 and 132GM constructs overall bound as well, or more efficiently than the atToc159-derived constructs to WT chloroplasts (Figure 9A); however, insertion of atToc132-derived constructs (atToc132, 132GM, 159A132GM) into WT chloroplasts was less efficient overall than atToc159-derived constructs as determined by ratios of inserted to bound receptor ($t = 2.79$, d.f. = 17, P = 0.01; Figure 9B), and insertion efficiency as a percentage of *in vitro* translation product added to the

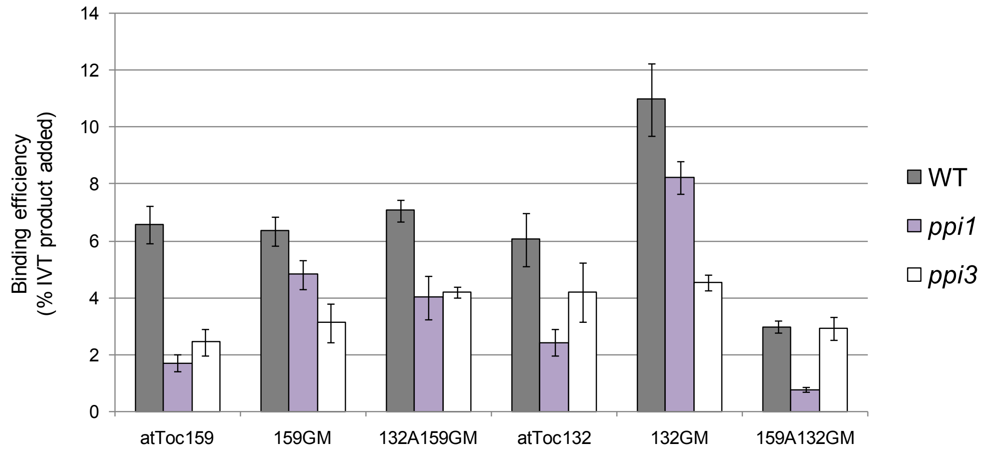
reaction (t = 6.35, d.f. = 44, $P < 0.001$; Figure 9C). This suggests that atToc159 is more efficiently inserted into the outer membrane – and therefore assembled into Toc complexes – of WT chloroplasts than atToc132 constructs *in vitro*. Removal of the A-domain of atToc159 resulted in a decrease in insertion into WT chloroplasts of ~50% when considering the ratio of inserted to bound receptor; however, this decrease was not found to be significant ($F_{(2,13)} = 1.85$, $P = 0.196$) (Figure 9B). Furthermore, when the insertion efficiency was calculated as a percentage of *in vitro* translation product added to the targeting reaction, no significant differences were found between atToc159-derived constructs ($F_{(2,21)} = 0.91$, $P = 0.42$) (Figure 9C).

2.4.2 Effect of the A-domain on targeting to structurally distinct Toc complexes

To assess if the A-domain may play a role in the specificity of targeting to isolated chloroplasts, binding and insertion levels of the constructs targeted to chloroplasts isolated from *ppi*1 (atToc33 knockout) and *ppi*3 (atToc34 knockout) were quantitated. Significant differences in binding of atToc159-derived constructs to *ppi*1 chloroplasts were observed ($F_{(2,15)} = 8.6$, $P = 0.003$), suggesting that the A-domain may have an effect on binding of atToc159 to atToc34-containing complexes. Removal of the A-domain of atToc159 resulted in an increase in binding to *ppi*1 chloroplasts relative to atToc159 (t = 5.31, d.f. = 8, $P < 0.001$) (Figure 10A). When the A-domain of atToc132 was added to the GM- portion of atToc159 (132A159GM), binding efficiency to *ppi*1 chloroplasts was also increased relative to atToc159 (t = 2.83, d.f. = 7, $P = 0.025$) (Figure 10A). The pattern of binding of the atToc132-derived constructs to *ppi*1 chloroplasts was similar to the pattern observed with WT chloroplasts – removal of the A-domain resulted in an increase in binding, while addition of the A-domain of atToc159 decreased binding. This may be indirect evidence that atToc132 preferentially interacts with atToc34, as in the absence of atToc33 (in the *ppi*1 mutant) the binding pattern of the atToc132-derived constructs is unchanged. The A-domain appeared to have little effect on binding to *ppi*3 chloroplasts as

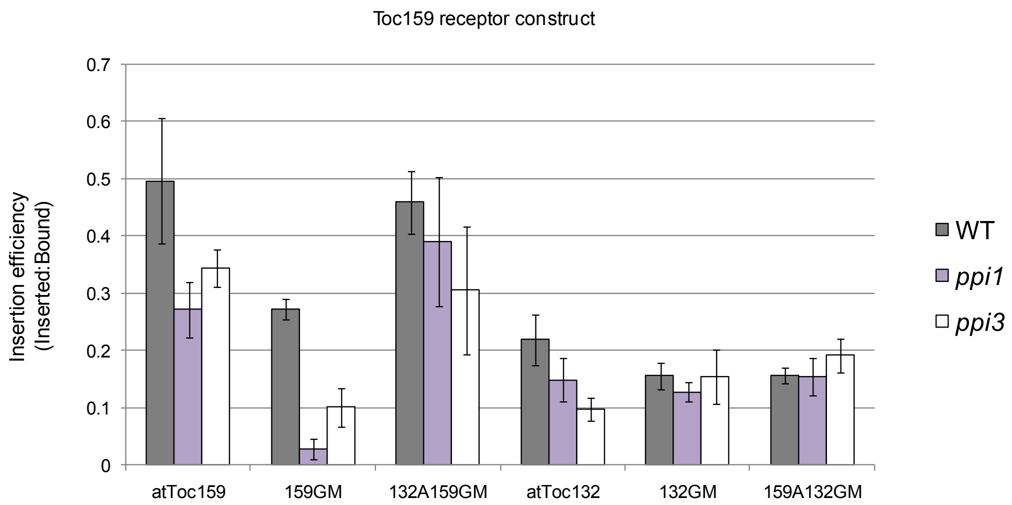
Figure 10. Targeting efficiency of Toc159 receptor constructs to chloroplasts isolated from WT, *ppi1* and *ppi3* Arabidopsis. Binding efficiency (% IVT product added) (A), inserted:bound (B) and insertion efficiency (% IVT product added) are shown. IVT product = *in vitro* translation product; error bars represent standard error of the mean.

A



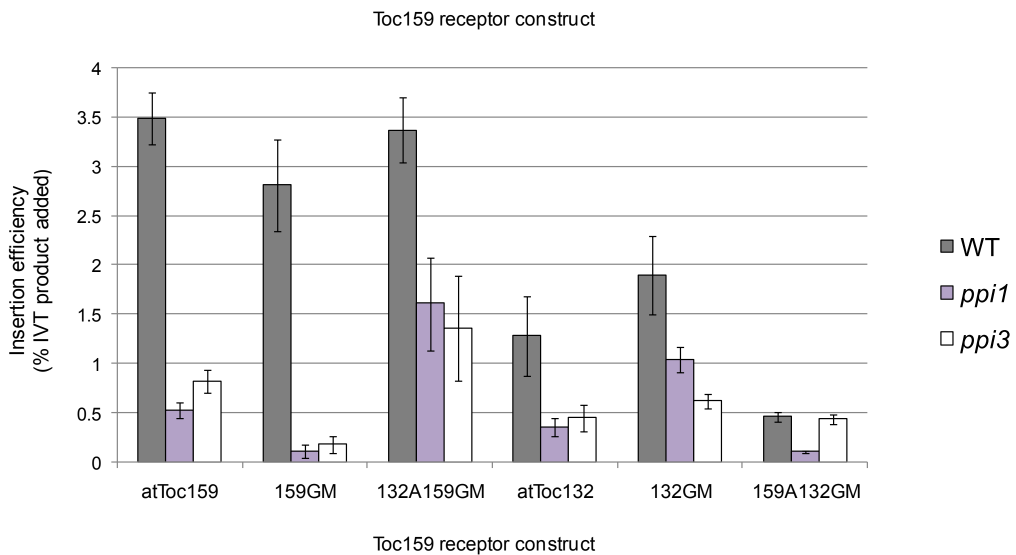
Toc159 receptor construct

B



Toc159 receptor construct

C



Toc159 receptor construct

evidenced by the relative lack of variation in binding efficiency among all constructs compared to binding to *ppi*1 chloroplasts (Figure 10A; SD_{ppi3} = 0.88; SD_{ppi1} = 2.69). It is unclear why differences in targeting to *ppi*1 were observed while variability among binding of contructs to *ppi*3 was minimal.

When looking at insertion, which is the determining step of Toc complex assembly, overall, no significant differences were observed among the atToc132-derived constructs and over all chloroplast types (Figure 10B; $F_{(8,46)} = 1.02$, $P = 0.43$). In contrast, the atToc159-derived constructs showed more variability in insertion levels. Specifically, a significant reduction in insertion of 159GM into *ppi*1 chloroplasts was observed relative to the insertion of atToc159 (Figure 10B; $t = 5.31$, d.f. = 8, $P < 0.001$). This suggests that removal of the A-domain of atToc159 resulted in the receptor binding more efficiently to atToc34-containing Toc complexes; however, subsequent insertion of 159GM at atToc34-containing complexes was not efficient. Binding of 159GM to *ppi*1 chloroplasts was approximately 3-fold more efficient than atToc159 (Figure 10A), but inserted:bound 159GM was approximately 5 times less than atToc159 (Figure 10B), indicating that the lower inserted:bound ratio can only partially be explained by high levels of initial binding. Addition of the A-domain of atToc132 onto 159GM (132A159GM) also had a stimulatory effect on insertion into atToc34-containing complexes of the *ppi*1 mutant, relative to atToc159 (Figure 10B; $t = 0.96$, d.f. = 7, $P = 0.37$).

2.5 Discussion

Previous studies have indicated that atToc159 is targeted to the Toc complex *via* a GTP-dependent interaction with atToc33, and requires the intrinsic GTPase activity of both receptors (Smith et al. 2002b; Bauer et al. 2002; Wallas et al. 2003). In addition to this homotypic interaction between the G-domains of the two proteins, insertion of atToc159 into the outer membrane also requires the presence of atToc75 – the protein channel in the chloroplast outer envelope membrane (Wallas et al. 2003). The A-domain of atToc159 is not essential for proper localization of the receptor to the chloroplast outer membrane as previously demonstrated by *in vitro* targeting assays using isolated chloroplasts (Smith et al. 2002b), and also by transient expression of a GFP-atToc159GM fusion protein in Arabidopsis protoplasts (Bauer et al. 2002). While the A-domain may not play a necessary role in targeting of atToc159 to the Toc complex, it remains to be determined if it may regulate the formation of structurally distinct Toc complexes. In this study, the role of the A-domain in targeting of atToc159 and atToc132 to chloroplasts isolated from wild type, *ppi*1 and *ppi*3 Arabidopsis was investigated to determine if the A-domain plays a regulatory role in the initial assembly of structurally (and functionally) distinct Toc complexes.

2.5.1 Effects of the A-domain on targeting of atToc159 and atToc132 to isolated chloroplasts

First, to determine if the A-domain may have a general effect on targeting of the Toc159 receptors to the Toc complex, A-domain truncated (132GM and 159GM) and A-domain swapped (132A159GM and 159A132GM) constructs were targeted to wild-type chloroplasts. Results demonstrated that atToc159 binding was not affected by removal of the A-domain or addition of the A-domain of atToc132. This suggests that the A-domain does not regulate binding of atToc159 to isolated chloroplasts, which is consistent with previous *in vitro* targeting experiments using wild type chloroplasts (Smith et al. 2002b). AtToc132 has been previously shown to target to isolated chloroplasts

(Bauer et al. 2000); however, the targeting characteristics of atToc132 have not been studied in detail. Since the G-domain of atToc132 is homologous to the G-domain of atToc159, it is likely that it also plays a central role in targeting of this receptor to the Toc complex, and in solid-phase binding assays, atToc132 was found to bind to both atToc33G and atToc34G (Ivanova et al. 2004). In the current study, binding of atToc132 was increased by the removal of its A-domain, and reduced by addition of the A-domain of atToc159; suggesting that the A-domain may have an inhibitory effect on binding of atToc132 to isolated chloroplasts. However, another interpretation may be that, removal of the A-domain results in an artificial decrease in steric hindrance imparted by the A-domain. The A-domain did not cause the same reduction in binding of atToc159 to WT chloroplasts. In vitro protein-protein interaction studies have demonstrated that the interaction between atToc132 and either atToc33 or atToc34 is of lower affinity than the interaction between atToc159 and atToc33 (Ivanova et al. 2004). Because of the low affinity interaction between atToc132 and either of the Toc34 homologues, binding of this receptor to isolated chloroplasts may be very sensitive to subtle structural differences with removal or swapping of the A-domain. In contrast, the high affinity interaction between the atToc159 and atToc33 may make differences in atToc159 binding caused by removal or swapping of its A-domain difficult to detect. While this explanation is speculative, it might suggest that binding levels to isolated chloroplasts are not necessarily representative of how the A-domain functions in targeting Toc159 receptors to chloroplasts *in vivo*.

Insertion of the Toc159 receptors into the outer membrane signifies complete assembly of the receptor into the Toc complex (Lee et al. 2003; Chen, Chen, and Schnell 2000) and is a requirement for successful preprotein import (Lee et al. 2003; Chen, Chen, and Schnell 2000). Targeting of Toc159 can be separated into two stages: binding and insertion. Insertion into isolated chloroplasts has been shown to be regulated by GTP, whereas initial binding of Toc159 to chloroplasts is energy-independent (Smith et al. 2002b). For these reasons, insertion of the Toc159 constructs into isolated chloroplasts in this study is a better indicator of the specificity of targeting of these receptors. The atToc132-derived

constructs were inserted into WT chloroplasts with a significantly lower efficiency than the atToc159-derived constructs (Figure 9B,C). These results are consistent with *in vitro* solid-phase binding assays that demonstrated that atToc132 had a relatively low affinity for the Toc34 homologues compared to the affinity of atToc159 for atToc33, given that atToc33 is the dominanat isoform present in green tissues of WT Arabidopsis (the source of chloroplasts used in this study). The observation that the A-domain has a more pronounced effect on binding of the atToc132-derived constructs than on insertion may be suggestive of a distinct mechanism of targeting of atToc132 from that used by atToc159. Studies on the energetics of targeting of atToc132 may reveal differences in mechanisms of targeting of atToc159 and atToc132.

Insertion of atToc159 into WT chloroplasts has previously been shown to be unaffected by removal of the A-domain (Smith et al. 2002b), and there were no significant differences in insertion of the atToc159-derived constructs into WT chloroplasts when insertion efficiency as a percentage of radiolabelled receptor added was considered (Figure 9C). Since the mechanism of the energy-independent binding step is not as well-characterized as insertion, comparing the inserted:bound ratio may provide more information about the specificity of Toc complex assembly. A slight decrease in inserted:bound atToc159 to WT chloroplasts with removal of the A-domain was observed (Figure 9B); however, this decrease was not statistically significant. Nevertheless, this small decrease may suggest that the A-domain may be important for efficient insertion of atToc159 into WT chloroplasts.

2.5.2 Role of the A-domain in targeting to structurally distinct Toc complexes

To investigate the role of the A-domain in targeting of the Toc159 receptors to structurally distinct Toc complexes, chloroplasts isolated from the *ppi*1 and *ppi*3 mutants were used for *in vitro* targeting assays. Of the atToc159-derived constructs, full-length atToc159 showed the lowest levels of binding to *ppi*1 chloroplasts, which may represent its lower affinity interaction with atToc34. Interestingly, removal of the A-domain resulted in an increase in binding to *ppi*1 chloroplasts. This

suggests that the A-domain of atToc159 may have an effect on binding by preventing or excluding an interaction with atToc34. This trend is also observed with binding of atToc132-derived constructs to WT and *ppi*1 chloroplasts. Again, it is unknown whether this exclusion is important for the mechanism of atToc159 binding, or if it is a result of a reduction in steric hindrance. Similarly, since the interaction between atToc159 and atToc34 is of lower affinity than the interaction between atToc159 and atToc33, it may be that binding to *ppi*1 chloroplasts is more sensitive to removal or addition of the A-domain compared to binding to WT chloroplasts. The ability of the 132A159GM construct to bind more efficiently than full-length atToc159 may be a result of a lack of specificity of the atToc132 A-domain for either atToc33 or atToc34.

No significant differences in inserted:bound receptor were observed between atToc132-derived constructs over all chloroplast types. This suggests that the A-domain does not play a role in the specificity of insertion of atToc132. In addition, the atToc132-derived constructs inserted with lower efficiency than atToc159 and 132A159GM, which is consistent with the hypothesis that the lower affinity interaction between atToc132 and atToc33G or atToc34G results in decreased insertion levels.

The most noticeable difference in insertion was the reduction of inserted:bound 159GM into *ppi*1 chloroplasts relative to full-length atToc159. This suggests that the A-domain may be important for insertion of atToc159 into the outer membrane, and may be especially important for its insertion into atToc34-containing complexes due to the inherently low affinity interaction between atToc159 and atToc34. Interestingly, insertion of 132A159GM into *ppi*1 chloroplasts is significantly higher than 159GM, suggesting that the A-domain of atToc132 is able to substitute for the A-domain of atToc159 in allowing insertion of this construct into the outer membrane. However, the inability of an A-domain to increase insertion of atToc132 indicates that both the A- and G- domains are together important for the specificity of insertion.

It is unknown how the A-domain may be involved in the insertion of the Toc159 receptors into the outer membrane. Insertion of atToc159 into isolated chloroplasts has been previously shown to be

stimulated by GTP hydrolysis. Therefore, it may be speculated that removal of the A-domain in these targeting assays may have an effect on the GTPase activity of either the Toc159 or Toc34 receptors. Recent structural studies on atToc33 and its pea orthologue, psToc34, have revealed a potential binding site for a co-GTPase-activating protein (coGAP) in the psToc34 homodimer complex that may be involved in stimulating GTP hydrolysis (Koenig, et al 2008). Therefore, it may be speculated that the A-domain could act as a coGAP, in the context of heterodimerization, stimulating GTP hydrolysis at Toc34 that would then theoretically result in insertion of Toc159 into the outer membrane. This is consistent with the observation that GTP hydrolysis activity of atToc33 is required for insertion of atToc159, and that GTP hydrolysis and insertion still occur in the absence of the A-domain.

2.5.3 Summary

In the current study, binding and insertion levels of theToc159 receptor constructs to isolated chloroplasts were used to assess a role for the A-domain in targeting of the Toc159 receptors to the Toc complex during initial Toc complex assembly. Toc complex assembly is completed when Toc159 receptors have been inserted into the outer membrane, therefore insertion is more important for the assembly of functionally distinct Toc complexes and may be more specific than the energy-independent binding step. Furthermore, binding may be influenced by multiple factors *in vivo*, making *in vitro* binding data difficult to interpret. Previous *in vitro* protein-protein interaction studies suggest the interaction between atToc159 and atToc33 is of higher affinity than 1) the interaction between atToc159 and atToc34, and 2) the interaction between atToc132 and either atToc33 or atToc34. In the targeting data presented in this study, insertion of the atToc159-derived constructs was overall more efficient than insertion of the atToc132-derived constructs. Given that atToc33 is the predominant Toc34 homologue present in green tissue, this may provide indirect confirmation of a preferential interaction between atToc159 and atToc33. Analysis of the insertion of atToc159-derived constructs into WT and *ppi*1 chloroplasts suggests that the A-domain of atToc159 is important for insertion of this

receptor into the outer membrane, especially into atToc34-containing complexes due to the relatively low affinity interaction between atToc159 and atToc34 compared to the interaction between atToc159 and atToc33. Furthermore, the A-domain of atToc132 is able to restore insertion of atToc159 to WT atToc159 levels, possibly due to the lack of specificity of the atToc132 A-domain for either atToc33 or atToc34.

In this study, targeting of the Toc159 receptors was examined in the context of initial Toc complex assembly. However, there exists the possibility that initial Toc complex assembly and preprotein targeting to the chloroplast do not occur independently. In the targeting assays used, chloroplast preproteins were not included. There is evidence to suggest that there may be a soluble pool of atToc159 in the cytosol, and this receptor may be able to cycle between a soluble and membrane-inserted form (Hiltbrunner et al. 2001). Although the mechanism of preprotein recognition by the Toc159 receptors is unknown, it is possible that atToc159 is able to recognize photosynthetic preproteins in the cytosol, and target them to the chloroplast *via* its highly specific interaction with atToc33. This highly specific pathway may act as a "funnel" for photosynthetic proteins when their expression is highly upregulated, allowing them to be quickly recognized in the cytosol, and efficiently targeted to the chloroplast. Furthermore, an interaction between the A-domain and precursor proteins may have an effect on whether atToc159 is part of the soluble pool, or the membrane-inserted form. However, the existence of a soluble pool of atToc159 remains controversial (Becker et al. 2004). The existence of a soluble pool of atToc132/120 has not been investigated, and this homologue is present at relatively low levels in green tissues. Therefore, if a soluble pool of atToc132/120 does exist, it may be difficult to detect. In addition, there are several lines of evidence to suggest that precursor proteins can stimulate GTP hydrolysis by the Toc complex (Jelic et al. 2002; Jelic, Soll, and Schleiff 2003; Schleiff, et al. 2003; Reddick et al. 2007). Because targeting of atToc159 is also regulated by GTP hydrolysis, this supports the hypothesis that targeting of atToc159 and precursor recognition may be linked.

The A-domain of atToc132 has little effect on insertion of this receptor into isolated chloroplasts, and insertion appears to be less efficient than insertion of atToc159. If atToc132 does not exist in a soluble pool, perhaps relatively less efficient insertion would be sufficient, as once inserted, the receptor may remain more stably associated with the Toc complex, prepared to import non-photosynthetic proteins. Alternatively, atToc132 may be inserted more efficiently in the presence of non-photosynthetic precursor proteins.

The use of chloroplasts isolated from the *ppi*1 and *ppi*3 mutants for *in vitro* targeting assays introduces several variables that cannot be controlled and may have an effect on levels of targeting of these receptors. For example, the relative levels of atToc34 and atToc33 in the *ppi*1 and *ppi*3 mutants, respectively, may be very different. AtToc34 is present at low levels in wild-type chloroplasts compared to atToc33. However, characterization of chloroplast protein import mutants has demonstrated that other components of the Toc complex may be upregulated in response to the mutation (Kubis et al. 2003). In addition, endogenous atToc159 and atToc132 associated with Toc complexes in isolated chloroplasts may also have an effect on the targeting characteristics of the radiolabelled Toc159 receptors. Evidence suggests that the stoichiometric ratio of Toc complex components is 3-4:3-4:1, Toc75:Toc34:Toc159 (Schleiff et al. 2003; Kikuchi, Hirohashi, and Nakai 2006). Therefore, initial binding of radiolabelled Toc159 receptor during *in vitro* targeting assays would be affected by pre-formed Toc complexes already containing Toc159. This is further evidence that the ratio of inserted:bound receptor may be a more accurate method of calculating insertion efficiency into isolated chloroplasts. This problem may be addressed by performing similar *in vitro* targeting assays using proteoliposomes, where the Toc complex composition is chemically defined. This approach has been used previously to investigate targeting properties of atToc159 (Wallas et al. 2003).

In conclusion, the data presented in this study provide evidence that the A-domain does play a role in targeting of the Toc159 homologues to structurally distinct Toc complexes. Information on the

structure of the A-domains presented in Chapter 2 of this study may lend insight into their role in insertion of the Toc159 homologues into the outer membrane during Toc complex assembly.

3. Structural analysis of the Toc159 family A-domains

3.1 Background

The Toc159 family of receptors have recently been proposed to belong to a growing class of intrinsically unstructured proteins (IUPs; Hernández Torres, Maldonado, and Chomilier 2007), which show lack of globular structure over their entire length or contain large unstructured regions. It is estimated that up to ~30% of all proteins in higher eukaryotes are intrinsically unstructured, or contain unstructured regions (Fink 2005). Several notable characteristics of the Toc159 family A-domains are consistent with their classification as intrinsically unstructured regions; for example, aberrant mobility during SDS-PAGE, high number of charged amino acid residues, presence of repetitive regions, and sensitivity to proteolysis (Tompa 2002; Dyson and Wright 2005). This proteolytic sensitivity is underscored by the initial identification of pea Toc159 as an ~86 kDa protein lacking most of the N-terminal A-domain (Hirsch et al. 1994; Kessler et al. 1994; Bölter, May, and Soll 1998; Chen, Chen, and Schnell 2000). IUPs have been implicated in a wide array of cellular activities including the regulation of transcription and translation, cellular signal transduction, post-translational modification, small molecule storage and the regulation of large multiprotein complex self-assembly (Dyson and Wright 2005). Their lack of structure is essential for their function which often involves coupled folding and ligand binding, interactions with multiple proteins, and participation in low affinity-high specificity interactions with their binding partners (Tompa and Fuxreiter 2007).

Several techniques are currently used to study IUPs including circular dichroism (CD) spectroscopy and fluorescence spectroscopy. CD spectroscopy takes advantage of the ability of chiral molecules, such as proteins, to absorb circularly polarized light (Berndt 1996). The difference in absorption of right- and left-circularly polarized light can give information about protein secondary structure. Far-UV CD spectra (spectra collected between 260 nm and ~180 nm) are sensitive to the conformation of the peptide backbone of proteins, and provide an estimate of the types of secondary structural elements present including α-helices, β-sheets, turns and unordered conformations, and can

also be used to detect overall conformational changes under different conditions (Berndt 1996). Figure 11 shows theoretical far-UV CD spectra of each type of secondary structure.

Fluorescence spectroscopy can use the instrinsic fluorescence of aromatic amino acids to detect structural changes within proteins. The aromatic amino acids tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe) have side chains that contain indole, phenol, and benzyl groups, respectively, that act as intrinsic fluorophores of proteins (Figure 12A; Lakowicz 2006). Trp has the strongest absorption and fluorescence characteristics relative to the other aromatic amino acids and unlike Phe and Tyr, Trp's fluorescence maximum shifts to shorter wavelengths in hydrophobic or constrained microenvironments (Figure 12B; Lakowicz 2006). Therefore, the fluorescence spectrum of Trp can be used to detect conformational changes that result in exposure of Trp's side chain to different microenvironments. This technique is useful for detecting structural changes within a protein under different experimental conditions.

3.2 Objectives and Hypotheses

In this study, CD and fluorescence spectroscopy were used to investigate the structure of the A-domain of atToc159 and atToc132. Based on previous indirect evidence, it was hypothesized that the A-domains would be unstructured under physiologically relevant conditions, and show conformational changes characteristic of IUPs at extreme temperature and pH.

Figure 11. Representation of typical far-UV CD Spectra associated with different types of secondary structure. Figure adapted from Mathews and van Holde (1996).

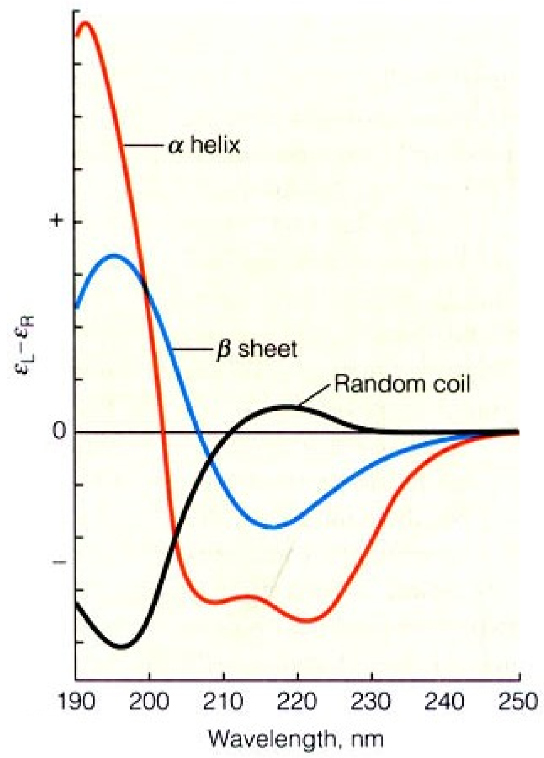
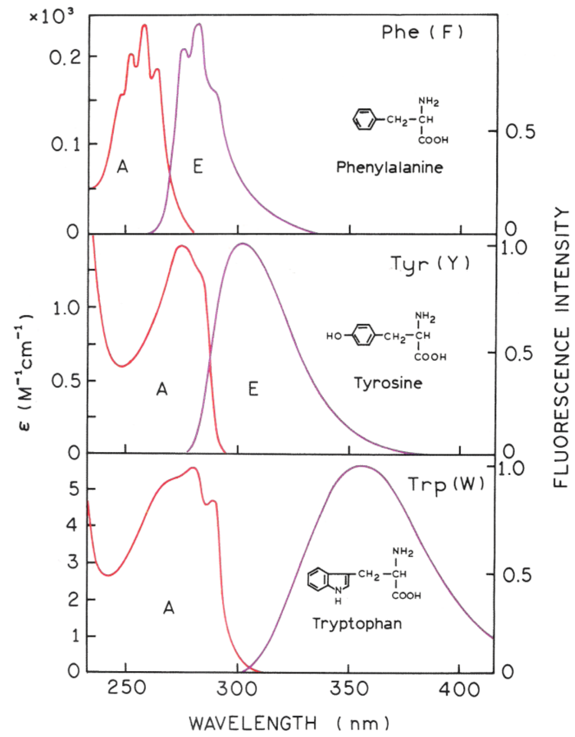
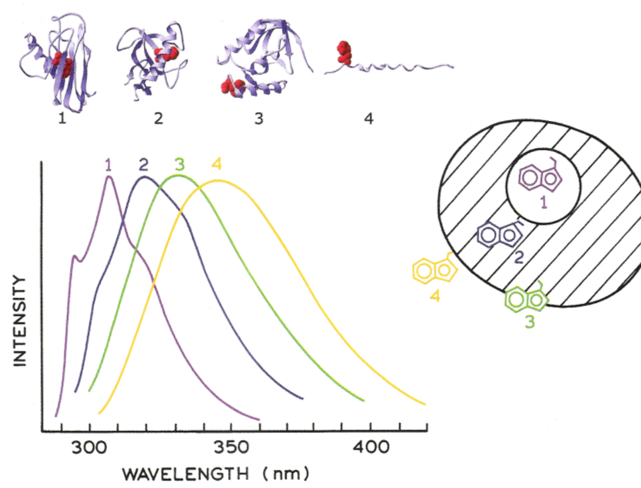

α helix
β sheet
Random coil
+
0
–
$\varepsilon_L - \varepsilon_R$
190
200
210
220
230
240
250
Wavelength, nm

Figure 12. Fluorescence characteristics of aromatic amino acid residues. The fluorescence absorption and emission spectra of the aromatic amino acid residues phenylalanine, tyrosine and tryptophan are shown in (A). Part (B) shows the effect of tryptophan environment on its emission spectra. Emission spectra 1-4 are of the proteins apoazurin Pfl, ribonuclease T_1, staphylococcal nuclease, and glucagon, respectively. Above the spectra are models of each protein with the single tryptophan residue in red. In the right hand schematic, the striped area represents a microenvironment where hydrogen bonding (i.e. with other residues or solvent) cannot occur. Figure taken from Lakowicz 2006.

A



B



3.3 Materials and methods

3.3.1 *Generation of* atTOC159A *and* atTOC132A *constructs*

Basepairs 1-2181 of *atTOC159* or 1-1365 of *atTOC132*, which correspond to the A-domain of atToc159 and atToc132, respectively (Ivanova et al. 2004), were sub-cloned into pET21b by a PCR-based method. The template used for amplification of *atTOC159A* was the *atTOC159* cDNA clone pET21d:atTOC159 (Smith et al. 2002b). The first round of PCR incorporated the coding sequence for 4 histidine residues at the 5' end of *atTOC159A*, and amplified the first 1096 bp of *atTOC159* using primer set 1 (appendix 3). The resultant ~1,100 bp fragment was then used as the template for a second round of PCR to incorporate the remainder of the sequence encoding the N-terminal 6 histidine residues and also to incorporate a 5' *NheI* restriction site for sub-cloning purposes *via* a second primer adapter (primer set 2, appendix 3). This fragment, coding for a 5' *Nhe I* restriction site, a 6xHis tag, and base pairs 1-1096 of *atTOC159* with incorporation of a 3' *KpnI* (introduced by a silent mutation within the primer), was blunt-end ligated into pCR4BluntTOPO according to the manufacturer's instructions (Invitrogen). Base pairs 1077-2181 of *atTOC159* were then amplified using primer set 3 (appendix 3), with incorporation of a *KpnI* site by a silent mutation at the 5' end, and blunt-end ligated into pCR4BluntTOPO according to the manufacturer's instructions (Invitrogen). Each of the fragments in pCR4BluntTOPO were digested with the appropriate restriction enzymes, and ligated into *NheI* and *SalI* restriction enzyme sites of pET21b in a triple ligation to generate pET21b:159A encoding the 159A protein with an N-terminal His-tag ($159A_{His}$). Positive transformants were confirmed by restriction digests and PCR.

The template used for amplification of *atTOC132A* was pET21a:*atTOC132* (Ivanova et al. 2004). Similar to atTOC159A subcloning, two rounds of PCR were used; the first for incorporation of a sequence corresponding to 4 His residues at the 5' end, and a *SacI* restriction site at the 3' end of *atTOC132A* (primer set 4, appendix 3). The resultant fragment was used as the template for a second

round of PCR, which incorporated the remainder of the sequence coding for 6 His residues as well as a 5’ *Nhe I* restriction site. This final fragment was blunt-end ligated into pCR4BluntTOPO and subsequently cloned into pET21b (Novagen) *via Nhe I* and *Sac I* restriction enzyme sites to generate pET21b:132A encoding the 132A protein with an N-terminal His-tag ($132A_{His}$).

3.3.2 Expression and purification of recombinant A-domains

pET21b:159A and pET21b:132A were transformed into *E. coli* strain BL21(DE3). $159A_{His}$ and $132A_{His}$ proteins were produced using the pET expression system (Novagen). Briefly, 1 L LB broth containing 50 µg/mL ampicillin was inoculated with a 1:100 dilution of overnight culture. Cultures were grown at 37°C, shaking at ~240 RPM until an OD_{600} of 0.6 – 0.8 was reached. Protein expression was induced with 1 mM IPTG and incubated at 37°C, shaking at 240 RPM for 3 h. Cultures were chilled on ice for 10 min and bacterial cells were harvested by centrifugation at 8,000 g for 10 min. The supernatant was discarded and the bacterial cell pellet was stored at -20°C until purification. Cell pellets were thawed on ice and resuspended in 20-30 mL of resuspension buffer (RB) (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 20 mM imidazole). Cells were lysed using a french press at a pressure of 18,000 psi, with a flow rate of approximately 20 drops per minute. The bacterial cell lysate was centrifuged at 50,000 g to pellet insoluble material.

Recombinant proteins were purified using immobilized metal affinity chromatography (IMAC). Briefly, a 1.5 mL column of Ni^{2+}-charged NTA resin (Novagen) was washed once with 6 column volumes of sterile milli-Q water, then twice with 4 mL of RB. The total soluble protein fractions were applied to the column twice followed by a wash with 6 column volumes of wash buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 30 mM imidazole) to remove any unbound or loosely bound protein. Proteins were eluted in 6 fractions of 750 µl elution buffer (EB) (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 250 mM Imidazole). Glycerol was added to a final concentration of 10% and protein samples were stored at -80°C until further use. $159A_{His}$ and $132A_{His}$ were further purified using a batch method

of ion exchange (Williams and Frasca 1998). Several IMAC elution fractions were combined to a final volume of 5-10 mL and diluted 1:1 in ion exchange binding buffer (IEBB; 20 mM piperazine pH 4.5, 0.2 M NaCl). The protein sample was then incubated with 1.5 mL of a strong-anion exchange resin in a glass vial (Q-sepharose Fast Flow ion exchange media, GE Health Sciences) for approximately 10 min at room temperature while rotating. The resin was allowed to settle, and the supernatant removed with a pasteur pipette. The resin was washed 3 times with 5 mL IEBB. Protein was eluted by incubation of the resin with 4 mL of ion exchange elution buffer (20 mM piperazine pH 4.5, 0.55 M NaCl) at room temperature while rotating. The resin was allowed to settle and the supernatant was transferred to a clean 15 mL screw-cap tube. The tube was centrifuged briefly on a benchtop centrifuge to pellet any residual resin, and the protein was transferred to a centrifugal filter device (Amicon Ultra-15 Ultracel 10k, Millipore) that had been equilibrated by addition of 14 mL of a CD-compatible buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl) followed by centrifugation at 4,000 g for 20 min. After addition of the eluted protein sample, 10 mL of CD buffer was added to the centrifugal filter to a final volume of ~14 mL. The filter was centrifuged at 4,000 g for 25 min, 20°C. The buffer was exchanged by two additional washes with 14 mL of CD buffer followed by centrifugation at 4,000 g for 25 min, 20°C. Proteins were visualized using SDS-PAGE with 4% stacking and 10% resolving gels stained with Coomassie Blue R250. Protein concentration was determined prior to CD spectroscopy using the Bio-Rad Protein Assay (Bio-Rad Laboratories Inc.).

3.3.3 Chemical modification of $132A_{His}$ and $159A_{His}$

Ion-exchange purified $132A_{His}$ and $159A_{His}$ in CD buffer were treated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) in the presence of excess ethanolamine as previously described (Graceffa, Jancsó, and Mabuchi 1992). The following components were added: 0.5 mg/mL protein in 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.2 M ethanolamine pH 6.0, 30 mM Mes pH 5.5 and 12 mM EDC. The reaction was incubated at room temperature for 3 hours and stopped by addition of

2x SDS-PAGE sample buffer. Proteins were visualized using Western blots ($132A_{His}$ and $159A_{His}$) and Coomassie blue staining (BSA). EDC modifies the carboxylic acid groups of glutamic and aspartic acid residues according to the following reaction (Graceffa, Jancsó, and Mabuchi 1992):

Protein-COO^- ⟶ protein-$CONHCH_2CH_2OH$

3.3.4 Western blots

$132A_{His}$ (~1 μg) and $159A_{His}$ (~10 ng) were resolved on 10% SDS-PAGE gels. Proteins were transferred to nitrocellulose membranes with a pore size of 0.2 μm (Protran, Schleicher and Schuell) using the semi-dry method. Briefly, following electrophoresis, the gel was soaked in transfer buffer (12.5 mM Tris, 96 mM glycine, 0.05% (w/v) SDS, and 10% (w/v) methanol) and transferred to nitrocellulose membranes using a Trans-Blot SD Semi-Dry transfer cell (Bio-Rad Laboratories Inc.) at 15 volts for 90 min. Following transfer, the nitrocellulose membrane was stored in TBS (20 mM Tris-HCl pH 7.5, 150 mM NaCl) overnight. The membrane was rinsed in water and stained with amido black (45% [v/v] methanol, 10% [v/v] acetic acid, 0.1% [w/v] amido black) to confirm successful transfer of proteins. The membrane was blocked by incubation with 5% (w/v) powdered milk in TBS-T (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20) at room temperature while shaking for 1 hour. The membrane was washed twice with TBS-T and incubated with primary antibody at room temperature, while rotating, for 2 hours. Primary antibodies used were rabbit antibodies raised against the A-domain of atToc159 (α-159A) diluted 1:2000 in TBS-T (1% BSA), and rabbit antibodies raised against the A-domain of atToc132 (α-atToc132A) diluted 1:5000 in TBS-T (1% BSA) (Ivanova et al. 2004, gift from D. Schnell, University of Massachusetts). After incubation with primary antibody, the membrane was rinsed 3 times in TBS-T for 5 min per rinse, and was subsequently incubated with secondary antibody (peroxidase conjugated goat anti-rabbit IgG (Rockland) diluted 1:5000 in TBS-T with 1% (w/v) BSA) at room temperature while rotating for 1 h. The membrane was washed five times with TBS-T for 5 min to remove any unbound secondary antibody. A chemiluminescence solution was

prepared containing 10 ml of 100 mM Tris-HCl, pH 8.5, 17µl of 2% (v/v) hydrogen peroxide, 25 µl Solution A (0.35 g of p-coumaric acid [4-hydroxycinnamic acid] dissolved in 25 ml of DMSO) and 50 µl of Solution B (1.1 g of luminol [5-Amino 2,3-dihydro 1,4-phtalazinedione] dissolved in 25 ml of DMSO). The solution was poured evenly over the surface of the nitrocellulose membrane, and incubated at room temperature for 4 min. Chemiluminescence signal on the membrane was detected using a Bio-Rad Fluor-S MultiImager in high sensitivity mode, using a Nikkor AF 50 mm lens (Nikon) with an f-stop of 1.4 and exposure time of 2 to 4 min. The membranes were analyzed using Quantity One 1-D Analysis software v4.6 (Bio-Rad Laboratories Inc.).

3.3.5 Circular Dichroism Spectroscopy

Far-UV CD spectra were measured on an Aviv 215 spectropolarimeter (Aviv Biomedical). Measurements were performed using rectangular quartz cells with 0.1 cm pathlength. $132A_{His}$ and $159A_{His}$ were typically measured at a concentration of 5 µM and 2.5 µM, respectively, in CD buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl), with added components where indicated, unless stated otherwise within figures. Samples were equilibrated at the indicated temperature for 10 min prior to measurements, and for pH-dependent studies the pH was measured prior to measurement. Spectra of protein samples and buffer alone were measured with a 0.5 nm/s scanning speed at 0.5 nm intervals, and are an average of four scans. Averaged buffer spectra were substracted from protein sample spectra and the resultant corrected spectra were smoothed and subsequently converted to mean residue ellipticity using Aviv CDSD software (Aviv Biomedical). Spectra were deconvoluted on the Dichroweb website (Whitmore and Wallace 2004) using the K2D method (Andrade et al. 1993).

3.3.6 Fluorescence Spectroscopy

Fluorescence spectra of $132A_{His}$ at varying pH, excited at 295 nm were measured on a Cary Eclipse fluorescence spectrophotometer (Varian, Palo Alto, CA, U.S.A.), using quartz cells of 1 cm pathlength. The slit widths for both excitation and emission wavelengths were 5 nm. Each spectrum is

an average of 10 scans at a scan rate of 600 nm/min with an averaging time of 0.1 s, and the corresponding average buffer spectrum was subtracted. Spectra of $132A_{His}$ at 2.5 μM in 10 mM Tris-HCl (of indicated pH), 50 mM NaCl, were measured at room temperature, and pH was confirmed prior to measurement.

3.4 Results

3.4.1 The A-domains of atToc159 and atToc132 are predicted to be natively disordered

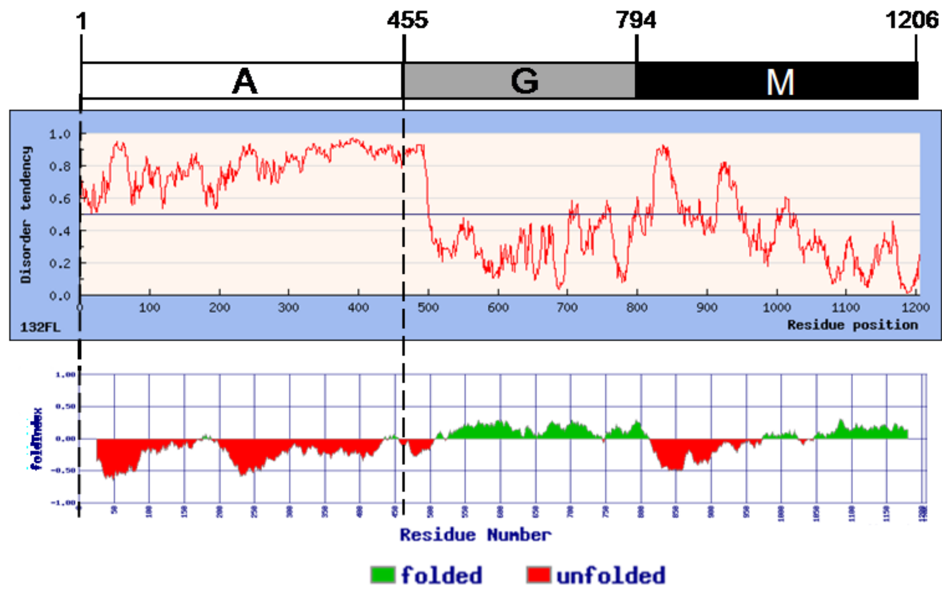
In this study, the A-domains were designated as amino acid residues 1-727 and 1-455, of atToc159 and atToc132, respectively, as previously described by Ivanova et al. (2004). The A-domain of atToc159 and atToc132 (159A and 132A, respectively) are rich in charged amino acid residues (~30%), and in particular contain an abundance of acidic residues (~25%), contributing to low theoretical pI values of 4.0 and 4.25 for 159A and 132A, respectively. The intrinsic disorder of each of atToc159 and atToc132 were predicted using IUPred (Dosztanyi et al. 2005a; Dosztanyi et al. 2005b) and FoldIndex (Prilusky et al. 2005). These prediction programs use experimental structural data obtained for both globular and unstructured proteins, as well as physical characteristics such as hydrophobicity and overall net charge to predict regions of disorder within proteins (Dosztanyi et al. 2005a; Dosztanyi et al. 2005b; Prilusky et al. 2005). These programs predicted the A-domain of both atToc132 and atToc159 to be mainly unfolded in comparison to the remainder of the corresponding full-length proteins at physiological pH (Figure 13 A,B).

3.4.2 Expression and purification of $132A_{His}$ and $159A_{His}$

E. coli-expressed His-tagged versions of 132A and 159A ($132A_{His}$ and $159A_{His}$) were subjected to immobilized metal affinity chromatography for purification (IMAC; Figure 14A, lane 2). To gain a level of purity suitable for CD spectroscopy, IMAC-purified protein was further purified by ion exchange (Figure 14A, lane 3). Expression of $132A_{His}$ and $159A_{His}$ was confirmed by Western blot analysis (Figure 14B). The theoretical molecular weights of $132A_{His}$ and $159A_{His}$ are approximately 50 kDa and 77 kDa, respectively; however, when these proteins are resolved using SDS-PAGE they migrate at an apparent molecular weight approximately 50 kDa larger than expected (Figure 14A). It has previously been shown that the full-length Toc159 receptors migrate at higher molecular weights than expected during SDS-PAGE (Bölter et al. 1998; Chen, Chen, and Schnell 2000); however, when

Figure 13. Disorder prediction of atToc132 and atToc159. Disorder within atToc132 (A) and atToc159 (B) were predicted using IUPred (top) and FoldIndex (bottom). Above the disorder predictions, a schematic representation of atToc132 and atToc159 are shown. Numbers above the schematics represent the amino acid number at the border between the domains indicated according to Ivanova et al. (2004). Dashed lines extend from the A-domain boundaries of the schematic to delineate the A-domain on the disorder prediction graphs.

A atToc132



B atToc159

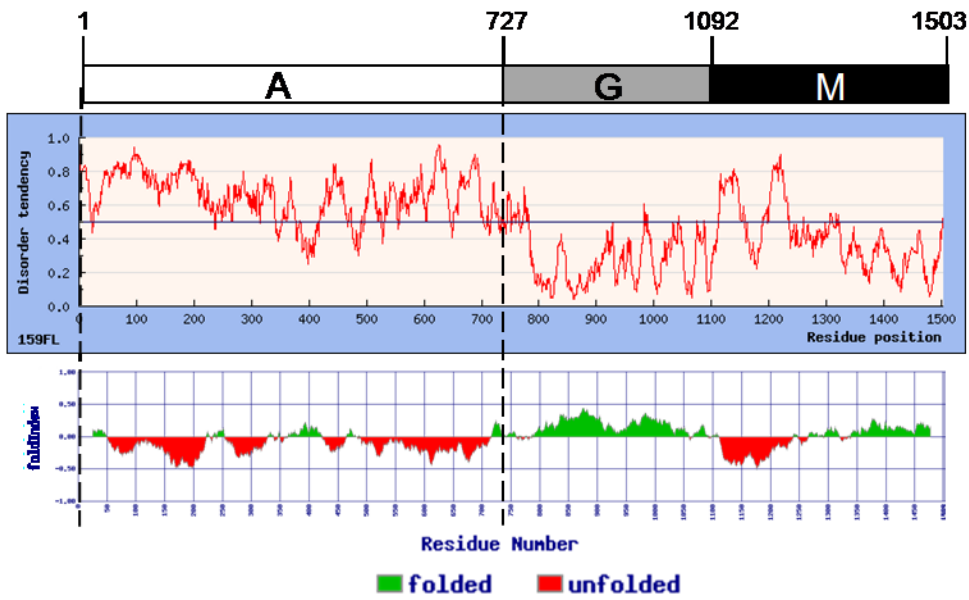
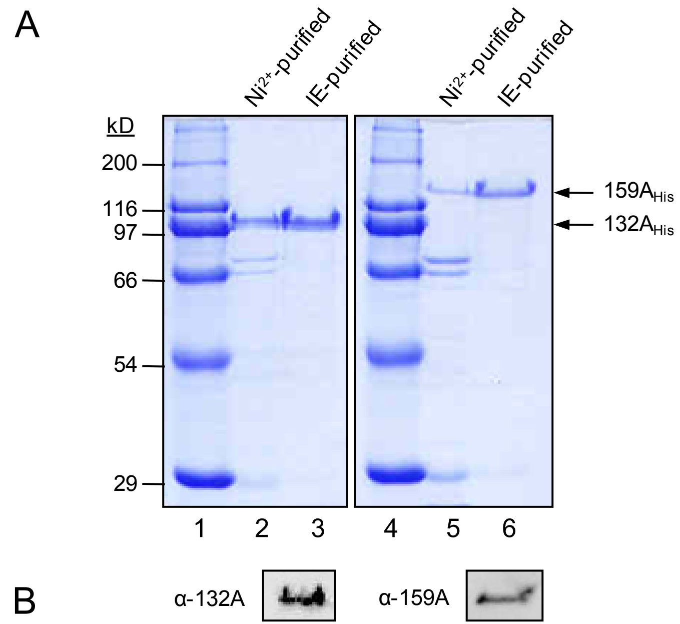


Figure 14. SDS-PAGE and Western blots of $132A_{His}$ and $159A_{His}$. (A) $132A_{His}$ and $159A_{His}$ were first purified using Ni^{2+}-NTA resin (lanes 2 and 5). To further purify these proteins from non-specifically bound proteins, $132A_{His}$ and $159A_{His}$ were subjected to ion exchange (lanes 3 and 6). Equal volumes of Ni^{2+}-purified and ion exchange (IE)-purified protein were loaded on the gel. The predicted molecular weights of $132A_{His}$ and $159A_{His}$ are 50 kD and 77 kD, respectively, demonstrating the aberrant migration of these proteins during SDS-PAGE. (B) Western blots were performed on ion exchange-purified $132A_{His}$ (1 µg) and $159A_{His}$ (10 ng) using antibodies raised against the A-domain of atToc132 and atToc159, respectively (Ivanova et al. 2004).


A
Ni2+-purified
IE-purified
Ni2+-purified
IE-purified
kD
200
116
97
66
54
29
159AHis
132AHis
1
2
3
4
5
6
B
α-132A
α-159A

the A-domain is degraded, they migrate as expected (Chen, Chen, and Schnell 2000), indirectly demonstrating aberrant electrophoretic mobility of the A-domains. Aberrant electrophoretic mobility is characteristic of acidic proteins, and it is thought to be caused by either abnormal binding of SDS or an abnormal shape of the SDS-protein complex (Graceffa, Jancsó, and Mabuchi 1992). To demonstrate that the anomalous electrophoretic migration of the A-domains is attributed to their high number of acidic amino acid residues, $132A_{His}$ and $159A_{His}$ were treated with EDC in the presence of an excess of the amine ethanolamine. EDC acts by converting negatively charged carboxylates into neutral amides (Hoare and Koshland 1967). In the presence of an excess of ethanolamine, glutamic and aspartic acid side chains are modified in such a way as to reduce their charge. Following EDC treatment, $132A_{His}$ and $159A_{His}$ migrated further in the gel, closer to their respective predicted molecular weights. Modification of each glutamic or aspartic amino acid side chain results in an increase in molecular weight of ~88 Da. $132A_{His}$ and $159A_{His}$ possess 115 and 198 potential modification sites, respectively, which corresponds to increased molecular weights of 60 and 95 kDa for $132A_{His}$ and $159A_{His}$, respectively. These numbers are very close to the observed molecular weights of modified $132A_{His}$ and $159A_{His}$ (Figure 15, lanes 2 and 4). In contrast, the migration of BSA during electrophoresis following EDC treatment did not noticeably change (Figure 15, lane 6). This demonstrates that the aberrant migration of the A-domains during SDS-PAGE is due to their abundance of acidic amino acid residues.

3.4.3 $132A_{His}$ and $159A_{His}$ are unstructured under physiological conditions

Circular dichroism (CD) spectroscopy was used to assess the secondary structure content of $132A_{His}$ and $159A_{His}$. In non-denaturing conditions (10 mM Tris-HCl pH 8.0, 50 mM NaCl) both $132A_{His}$ and $159A_{His}$ show far-UV spectra typical of natively unfolded proteins, which are characterized by the presence of a deep minimum in the vicinity of 200 nm and a relatively low ellipticity at ~220 nm (Uversky 2002a; Figure 16). Far-UV CD spectra were deconvoluted using the K2D method (Andrade et al. 1993) on the DICHROWEB website (Whitmore and Wallace 2004). Using this method it was

Figure 15. EDC treatment of $132A_{His}$ and $159A_{His}$. Modification of $132A_{His}$ and $159A_{His}$ with EDC in an excess of ethanolamine resulted in a molecular weight shift during SDS-PAGE (Compare lanes 1 and 2, 3 and 4), whereas no molecular weight shift was observed for BSA (compare lanes 5 and 6). $132A_{His}$ and $159A_{His}$ were detected using Western blotting, BSA was stained with Coomassie blue.

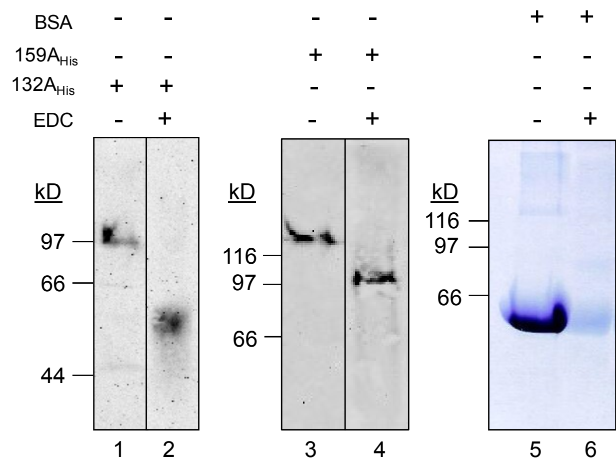
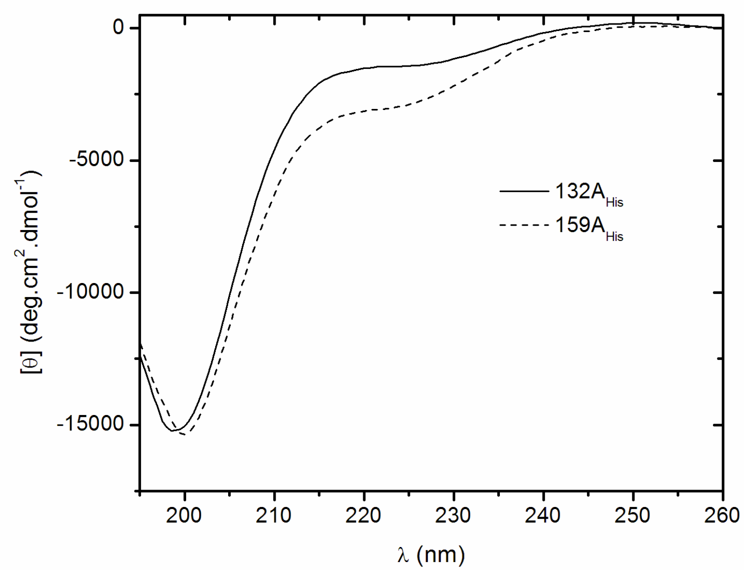

BSA
159A
His
132A
His
EDC
kD
97
66
44
kD
116
97
66
kD
116
97
66
1
2
3
4
5
6

Figure 16. CD spectra of $132A_{His}$ (5 μM) and $159A_{His}$ (5 μM) at 25°C, pH 8. Spectra demonstrate that $132A_{His}$ and $159A_{His}$ are mainly random coil at 25°C and physiological pH, with some residual secondary structure.


0
-5000
-10000
-15000
[θ] (deg.cm2.dmol-1)
132AHis
159AHis
200
210
220
230
240
250
260
λ (nm)

estimated that 73% and 63% of residues in $132A_{His}$ and $159A_{His}$, respectively, are involved in random coil secondary structure. This indicates that at physiological temperature and pH, the A-domains are mainly unstructured and supports the hypothesis that these proteins are IUP domains.

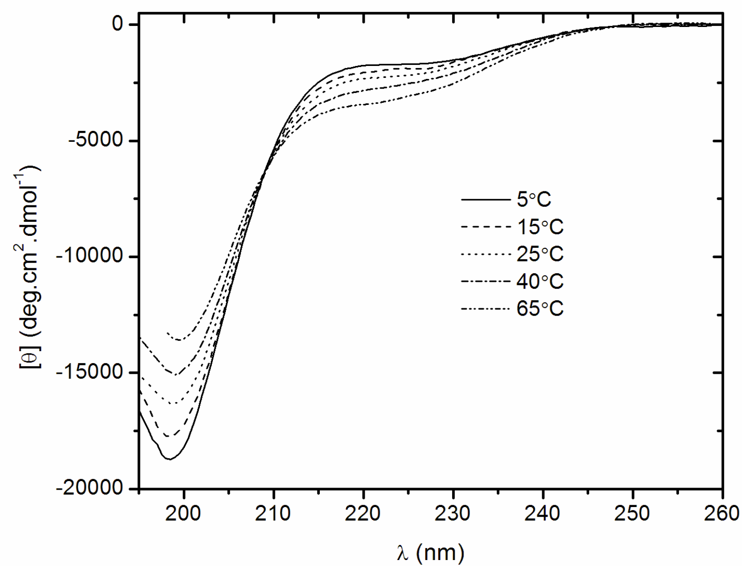
3.4.4 Temperature and pH dependent structural changes in $132A_{His}$ and $159A_{His}$

To further characterize the structural properties of the A-domains, the effects of temperature and pH on the conformation of $132A_{His}$ and $159A_{His}$ were investigated. Both $132A_{His}$ and $159A_{His}$ exhibit a temperature-induced gain in secondary structure, evidenced by a decrease in ellipticity at ~220 nm with increasing temperature (Figure 17). This modest, linear gain in secondary structure is characteristic of IUPs (Uversky 2002a), and is in contrast to the sharp loss of structure associated with the heating of globular proteins.

The effects of pH on the structure of $132A_{His}$ and $159A_{His}$ were also investigated. $132A_{His}$ exhibited only small differences in far-UV CD spectra observed at pH 8 and 10 (Figure 18A). However, this protein showed a substantial increase in secondary structure at pH 3. The intrinsic fluorescence of $132A_{His}$ at low and neutral pH supports this observation; fluorescence of $132A_{His}$ excited at 295 nm shows a shift in its maximum to a higher wavelength with increasing pH, indicating that at low pH, Trp residues are less exposed to the solvent, whereas at neutral pH, Trp residues become more exposed (Figure 18B). $132A_{His}$ possesses 2 Trp residues, both centrally located within the protein (residues 225 and 234, respectively). Completely exposed Trp has a maximum fluorescence around 350 nm. Based on the intrinsic fluorescence of $132A_{His}$ at 350 nm, it was demonstrated that the Trp residues of $132A_{His}$ are largely exposed at neutral pH. This is consistent with the far-UV CD spectra, which demonstrate $132A_{His}$ to be unstructured at neutral pH (Figure 17A). The fluorescence spectra of $159A_{His}$ at low and neutral pH were not measured, as $159A_{His}$ does not possess any Trp residues. However, far-UV CD spectra of $159A_{His}$ indicate that this protein also gains secondary structure at low pH (Figure 18C). This increase in secondary structure of $132A_{His}$ and $159A_{His}$ at low pH can be attributed to an overall

Figure 17. CD spectra of $132A_{His}$ and $159A_{His}$ with changing temperature. CD spectra of (A) $132A_{His}$ and (B) $159A_{His}$ were measured over increasing temperature. Changes in mean residue ellipticity at ~198 nm and ~222 nm indicate that with increasing temperature, these proteins show an increase in secondary structure.

A



B

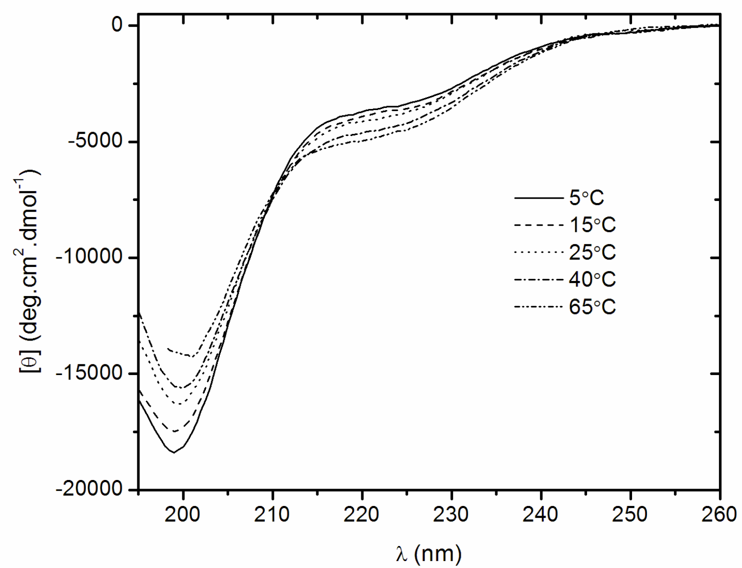
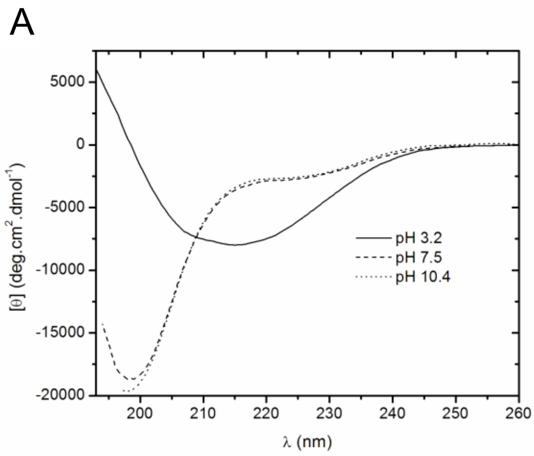
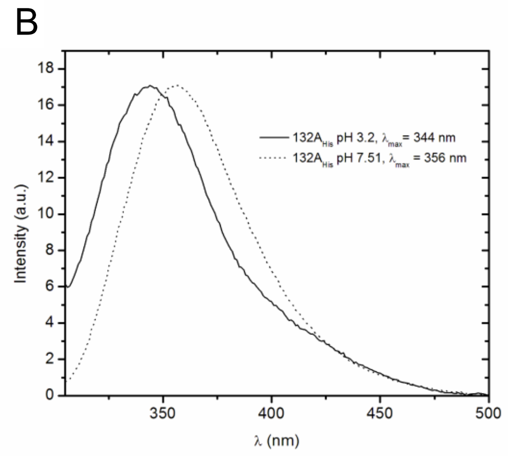
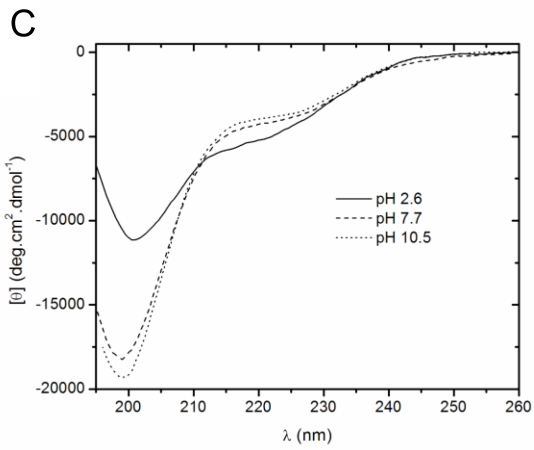


Figure 18. pH-dependent structural changes in $132A_{His}$ and $159A_{His}$. Both $132A_{His}$ and $159A_{His}$ show an increase in secondary structure at low pH as shown by their far-UV CD spectra (A and C, respectively.) In addition, a shift in the fluorescence emission spectrum of $132A_{His}$ (295 nm excitation wavelength) shows that its Trp residues become more exposed at neutral pH (B). The far-UV CD spectra of $159A_{His}$ collected at varying pH (C) also shows that this protein contains more secondary structure at low pH. The fluorescence spectra of $159A_{His}$ at low and neutral pH was not measured, as $159A_{His}$ does not contain any Trp residues.


A
5000
0
-5000
-10000
-15000
-20000
200
210
220
230
240
250
260
[θ] (deg.cm^2.dmol^{-1})
λ (nm)
pH 3.2
pH 7.5
pH 10.4


B
18
16
14
12
10
8
6
4
2
0
350
400
450
500
Intensity (a.u.)
λ (nm)
132A$_{His}$ pH 3.2, λ_{max}= 344 nm
132A$_{His}$ pH 7.51, λ_{max} = 356 nm


C
0
-5000
-10000
-15000
-20000
200
210
220
230
240
250
260
[θ] (deg.cm^2.dmol^{-1})
λ (nm)
pH 2.6
pH 7.7
pH 10.5

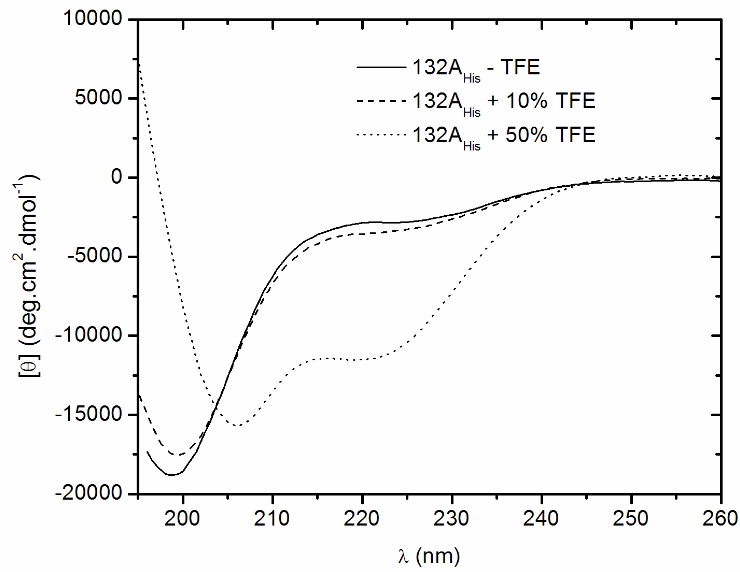
decrease in net charge at pH 3, which is below the theoretical pI values of both $132A_{His}$ and $159A_{His}$ (4.25 and 4.0, respectively). A decrease in net charge leads to a decrease in electrostatic repulsion between residues, allowing partial folding of the protein.

3.4.5 Structure of $132A_{His}$ and $159A_{His}$ in the presence of TFE

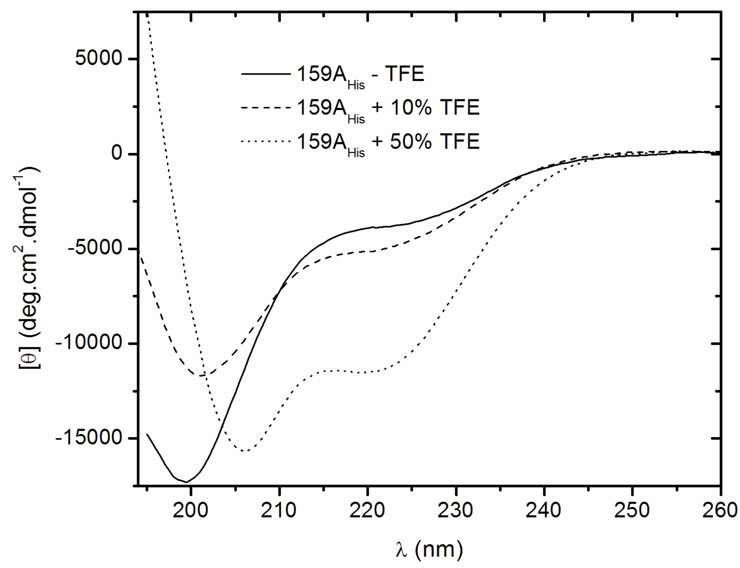
Far-UV CD spectra of the A-domains in the presence of trifluoroethanol (TFE) revealed a considerable increase in secondary structure (Figure 19). $159A_{His}$ appeared to gain more structure in 10% TFE compared to $132A_{His}$, suggesting that the A-domain of atToc159 may be relatively more sensitive to conformational changes. In 50% TFE $132A_{His}$ and $159A_{His}$ showed similar spectra, with two minima at approximately 220 nm and 208 nm characteristic of alpha-helices. Deconvolution of A-domain spectra in the absence of TFE and in 50% TFE revealed an increase in alpha helical structure from about 4% to 28% for both proteins. The physiological significance of TFE-induced secondary structure remains controversial; however, the ability of the A-domain to gain structure in the presence of TFE may suggest a propensity for structure under suitable conditions (Receveur-Bréchot et al. 2006).

Figure 19. CD spectra of $132A_{His}$ and $159A_{His}$ in Buffer/TFE mixtures. $132A_{His}$ (A) and $159A_{His}$ (B) were prepared in 10% or 50% mixtures of TFE. Both proteins show an increase in secondary structure in the presence of TFE.

A



B



3.5 Discussion

Several observations led to the hypothesis that the A-domain is intrinsically unstructured, including its abundance of acidic amino acid residues, aberrant mobility during SDS-PAGE and observed proteolytic sensitivity. This proteolytic sensitivity is underscored by the initial identification of pea Toc159 as an 86 kDa protein, which was later found to be a fragment resulting from proteolytic degradation of the A-domain (Bölter, May, and Soll 1998; Chen, Chen, and Schnell 2000). Both atToc159 and atToc132 were predicted to be mainly unstructured using IUPred (Dosztanyi et al. 2005a; Dosztanyi et al. 2005b) and FoldIndex (Prilusky et al. 2005) (Figure 13). Furthermore, hydrophobic cluster analysis (HCA) of the Toc GTPases has recently suggested that the A-domain of atToc159 evolved to become unstructured (Hernández Torres, Maldonado, and Chomilier 2007). Together, these observations support the hypothesis that the A-domains of atToc159 and atToc132, respectively, are intrinsically unstructured.

In this study, CD spectroscopy and fluorescence spectroscopy were used to investigate the structure of recombinant, His-tagged versions of the A-domain of both atToc159 and atToc132. Under physiological conditions, the A-domains were found to be mainly unstructured (Figure 16). Furthermore, temperature-induced folding (Figure 17) and increases in secondary structure at low pH (Figure 18) were observed for both $132A_{His}$ and $159A_{His}$, which are characteristics of intrinsically unstructured proteins (Uversky 2002a). Alpha helical content of $132A_{His}$ and $159A_{His}$ was considerably increased in the presence of 50% trifluoroethanol (TFE) (Figure 19). TFE is known to induce structure in proteins by displacing water molecules surrounding the protein thereby decreasing hydrogen bonding with water molecules, resulting in more inter-residue interactions (Roccatano et al. 2002). This displacement of water molecules is thought to mimic certain *in vivo* conditions; for example, the interface between the surface of a protein molecule and a biological membrane or binding partner. The structural changes observed over increasing temperature, low pH and in the presence of TFE are all

related to the physical characteristics of IUPs. Structure is induced either by an increase in hydrophobicity or by a reduction in electrostatic repulsion, promoting intramolecular interactions. *In vivo*, binding of IUPs to specific target molecules also influences hydrophobicity and overall net charge which is critical for their inter-molecular interactions. Complexes of several IUPs with their ligands have net charges and hydrophobicities comparable to globular proteins (Uversky 2002a).

The unique physical properties of IUPs give them distinct functions such as the ability to engage in low affinity, high specificity interactions, an increased speed of interaction and the ability to bind several different partners (Andrade, Perez-Iratxeta, and Ponting 2001; Tompa 2002; Dyson and Wright 2005; Fink 2005). It has previously been suggested that the repetitive nature of the A-domain may be indicative of protein-protein interactions (Chen, Chen, and Schnell 2000). However interaction of the A-domain with other proteins has not been documented. The general properties of IUPs, as well as the identification of Toc159 as the primary preprotein receptor make the A-domain an ideal candidate for binding transit peptides. Transit peptides show a lack of sequence conservation and are highly variable in length (Bruce 2001), and the receptors of the Toc complex must be able to recognize these highly diverse targeting signals. The flexible nature of the unstructured A-domain may allow this receptor to recognize transit peptides that are highly variable in primary sequence. The ability of the A-domain to bind transit peptides with low affinity would allow the precursor protein to be easily passed to additional potential transit peptide binding sites on Toc159, or on Toc34 and Toc75. Indeed, preproteins have been shown to interact directly with Toc34 and Toc75 during protein import (Schnell, Kessler, and Blobel 1994). This is reminiscent of how mitochondrial protein transport is proposed to occur – with precursors being guided across the mitochondrial membranes *via* a chain of protein binding events, ensuring unidirectional import (Pfanner 2000). While the affinity of one receptor for preproteins may be low, an interaction with several components provides the specificity for preproteins required to maintain the fidelity of import. A low affinity interaction with transit peptides may explain why an interaction with the A-domain is undetectable using traditional binding assays or cross-linking

approaches, and may also explain why reports of the order of preprotein binding to Toc34 and Toc159 have been conflicting (Ma et al. 1996; Schleiff, Jelic, and Soll 2003). In addition, binding to preproteins may induce conformational changes either within the A-domain or within the preprotein that are important for triggering subsequent stages in import. Interestingly, it has recently been shown that translocation of mitochondrial inner-membrane proteins lacking a presequence involves the recognition of conformational signals by natively unfolded regions found within receptor components of the **t**ranslocon at the **o**uter membrane of **m**itochondria (Tom), including Tom70 and Tom22 (Marcos-Lousa, Sideris, and Tokatlidis 2006).

When entered into the ELM (Eukaryotic Linear Motif) server (Puntervoll et al. 2003) several putative linear motifs are identified within the A-domain of both atToc132 and atToc159 including a 14-3-3 protein binding site and multiple putative phosphorylation sites. This prediction server does point out that short patterns applied to proteins are usually not statistically significant, and produces a large number of false positives; however, it may suggest other functions, or modes of regulation of the A-domain. It has been shown that both Toc34 and Toc159 are able to be phosphorylated *in vitro* (Fulgosi and Soll 2002); though, the potential effects of receptor phosphorylation on import remain unclear (Aronsson et al. 2006). It has also been suggested that phosphorylated preproteins may be associated with a guidance complex consisting of a 14-3-3 protein dimer and cytosolic Hsp70 (May and Soll 2000) that increases the efficiency of import (May and Soll 2000). Interestingly, 14-3-3 proteins possess a "clamp-like" dimeric structure and are well-known as mediators of protein-protein interactions (Finnie, Borch, and Collinge 1999). Therefore it is possible that the A-domain may interact with a preprotein guidance complex containing a 14-3-3 protein, which allows transfer of the preprotein from the guidance complex to Toc159. An interaction between the Toc159 receptors and the putative guidance complex has not been shown experimentally.

3.5.1 Potential role for the A-domain in the functional diversity within the Toc159 family

Genetic and biochemical studies have revealed that the members of the Toc159 family in Arabidopsis are functionally distinct – atToc159 is involved in the import of photosynthetic proteins and atToc132 and atToc120 are functionally redundant, importing non-photosynthetic or plastid housekeeping proteins (Bauer et al. 2000; Ivanova et al. 2004). It has been proposed that, due to its diversity in primary structure across family members, the A-domain is responsible for the functional specificity of these receptors (Bauer et al. 2000).

It has recently been proposed using hydrophobic cluster analysis (HCA) that the A-domain of atToc159 evolved from duplication of an ancient G-domain also of ancestry to pea Toc34 and the G-domain of atToc159 (Hernández Torres, Maldonado, and Chomilier 2007). The authors suggest that during evolution, these repetitive regions lost their overall fold to become intrinsically unstructured domains (Hernández Torres, Maldonado, and Chomilier 2007). Interestingly, a common feature of IUPs is their evolution by repeat expansion (Tompa 2003). Studies on the evolution of protein repeats have revealed the most common function resulting from repeat expansion within proteins is protein binding (Andrade et al. 2001). Investigation of several IUPs has revealed that repetitive segments are essential for the function of IUPs and fall into 3 major functional classes. Type I repeats retain their function (i.e. bind the same ligand), leading to several regions with the same function. Type II repeats may acquire new functions (i.e. bind new substrates) allowing IUPs to interact with multiple, different binding partners. Finally, type III repeats result in a physiologically novel function when enough repeats arise (Tompa 2003). It has been previously suggested that differences in size and overall net charge of the A-domains may reflect their functional specificities by binding to transit peptides with varying affinities (Bauer et al. 2000). Mutant analysis of the transit peptide of the small subunit of Rubisco has revealed an importance for the overall context of the transit peptide, and several regions of the transit peptide have been found to be functionally redundant (Lee et al. 2006). Therefore, it may be hypothesized that

the A-domain of atToc159 and one of its preferred substrates, pSSU (a photosynthetic precursor protein), have acquired repeats that mediate their specific binding to each other. This may have happened with other photosynthetic preprotein transit peptides as well. In this scenario, repeat expansion of the A-domain of atToc159 may have conferred more specificity of this receptor for photosynthetic proteins in comparison to the specificity of atToc132 and atToc120 for non-photosynthetic proteins. It would be interesting to analyze transit peptides of preproteins of distinct functional classes (i.e. photosynthetic versus non-photosynthetic) to measure differences such as overall length or charge. Differences in length or the presence of repeats in transit peptides of preproteins belonging to different functional classes may provide information about potential specific interactions between transit peptides and the A-domains.

Alternatively, transit peptide binding may be a novel function possessed by atToc159, and not other members of the Toc159 family, as a result of more extensive repeat expansion. This may be theoretically possible, as the A-domain is not required for protein import (Lee et al. 2003; Chen, Chen, and Schnell 2000). In this case, the mechanism of preprotein recognition of the atToc132/120 subgroup of receptors may be distinct from that of atToc159. Hernández Torres et al (2007) identify repeated regions within the A-domain of atToc159. It would be interesting to conduct structural analyses of each of these repetitive domains and investigate interactions with transit peptides using highly sensitive biophysical techniques such as isothermal titration calorimetry that are able to detect low affinity interactions. Also of note, of the repeated regions in atToc159 identified by Hernández Torres et al. (2007), the region in closest proximity to the G-domain showed the highest homology to other Toc159-related proteins including atToc132 and atToc120, as well as putative Toc159 homologues from other species. The relatively high sequence similarity in this region among the identified homologues may suggest that this region has a similar function across the receptors, and additional repeats N-terminal to this region may dictate the functional differences between these receptors. For example, additional repeats in the A-domain of atToc159 may confer additional, novel functions. It has also been suggested

that protein-protein interactions mediated by unstructured proteins are somewhat independent of their sequence, and that it is the physical properties of the unstructured region that is important for its function (Tompa and Fuxreiter 2007). Therefore, the possibility does exist that the A-domains of the Toc159 family of receptors are not responsible for the specificity of these receptors.

In summary, the classification of the A-domain as an intrinsically unstructured region of the Toc159 receptors is important when considering the preprotein binding properties of this receptor. While shown to not be essential for import, *in vivo* the A-domains may represent the first point of contact between preproteins and the Toc complex. Such an interaction may be bypassed *in vitro*, explaining why a function for the A-domain remains elusive.

4. Conclusions

Chapter 2 of this study demonstrates that the A-domains of atToc132 and atToc159 are unstructured under physiological conditions, and likely belong to the class of intrinsically unstructured proteins. As proteins of this class are commonly involved in protein-protein interactions, often with multiple binding partners, this structural characteristic allows for speculation about potential function(s) of the A-domains in chloroplast protein import. Although interaction of the A-domain with other proteins has not been demonstrated experimentally, it may be proposed that the intrinsically unstructured A-domains may play a role in preprotein recognition by binding transit peptides.

Chapter 1 of this study was aimed at studying the effects of the A-domain on targeting of the Toc159 receptors to structurally distinct Toc complexes. It was hypothesized that the A-domain contributes to the formation of structurally distinct Toc complexes due to their unique amino acid sequences. The results suggested that the A-domain does have an effect on targeting of atToc132 and atToc159 to the Toc complex, but that the two receptors may employ slightly different mechanisms of targeting. The knowledge that the A-domains are intrinsically unstructured suggests their involvement in low affinity interactions (a hallmark of IUPs) with other potential binding partners (e.g. other components of the Toc complex). Such low-affinity interactions may preclude an ability to detect specific effects of this domain on binding to isolated chloroplasts in the *in vitro* system used in this study. For example, it is already known that atToc33 and atToc34 have redundant functions and are each able to engage atToc159 and atToc132/120 (Ivanova et al. 2004). Therefore, low affinity interactions between the A-domain and the Toc34 receptors that result in specific binding *in vivo*, may be undetectable *in vitro* using isolated chloroplasts.

The *in vitro* targeting data also seem to suggest a role for the A-domain in insertion of atToc159 into isolated chloroplasts. It may be that once atToc159 is bound to the outer membrane in association with the Toc complex, the A-domain interacts with other Toc complex components leading to a conformational change that is more conducive to insertion of atToc159 into the outer membrane. IUPs

commonly undergo induced folding, coupled with binding to their interacting partner(s) (Fuxreiter et al. 2004), and it has been shown in mitochondrial protein import that conformational changes in components of the import translocon act as signals for protein import. Therefore a potential conformational change induced in atToc159 by interaction of the A-domain with other Toc complex components may act as a signal to induce this receptor's insertion. The observation that atToc132 insertion is not affected by the A-domain may suggest that insertion of atToc159 may have become specialized over evolution in order to maximize its efficient assembly into the Toc complex. Efficient assembly of atToc159-containing complexes would accommodate the vast number of photosynthetic proteins that are expressed and need to be imported during photomorphogenesis. The A-domain of atToc159 may also be optimized to allow a soluble pool of atToc159 (Hiltbrunner et al. 2001) to be more easily targeted to the Toc complex. In contrast, the need for efficient insertion may not be as great for atToc132 (and atToc120) if it exists as a more stable part of the Toc complex. It is interesting to speculate that the difference in size of the A-domains of atToc159 and atToc132/120 may reflect differences in targeting mechanisms; the additional repeats in the A-domain of atToc159 may allow this receptor to associate simultaneously with transit peptides and atToc33 at the Toc complex.

Future research should be aimed at identifying putative binding partners of the A-domain (for example, transit peptides, and components of the Toc complex or otherwise). This may clarify the *in vitro* targeting assay data collected in this study, and will also be important for understanding several aspects of protein import that remain unclear including preprotein recognition and Toc complex assembly.

References

Andrade, M.A., P. Chacón, J.J. Merelo, and F. Morán. 1993. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot. Eng.* 6: 383-390.

Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting. 2001. Protein repeats: Structures, functions and evolution. *J. Struct. Biol.* 134: 117-131.

Aronsson, H., J. Combe, R. Patel, and P. Jarvis. 2006. *In vivo* assessment of the significance of phosphorylation of the Arabidopsis chloroplast protein import receptor atToc33. *FEBS Lett.* 580: 649-655.

Arnon, D.I. 1949. Copper enzymes in isolated chloroplasts in *Beta vugaris*. *Plant Phys.* 24: 1-15.

Bae, W., Y.J. Lee, D.H. Kim, J. Lee, S. Kim, E.J. Sohn, and I. Hwang. 2008. AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. *Nature Cell Biol*. 10: 220-227.

Bauer, J., K. Chen, A. Hiltbrunner, E. Wehrli, M. Eugster, D.J. Schnell, and F. Kessler. 2000. The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature* 403: 203-207.

Bauer, J., A. Hiltbrunner, P. Weibel, P.A. Vidi, M. Alvarez-Huerta, M.D. Smith, D.J. Schnell, and F. Kessler. 2002. Essential role of the G-domain in targeting of the protein import receptor atToc159 to the chloroplast outer membrane. *J. Cell Biol.* 159: 845-854.

Becker, T., M. Jelic, A. Vojta, A. Radunz, J. Soll, and E. Schleiff. 2004. Preprotein recognition by the Toc complex. *EMBO J.* 23: 520-530.

Bedard, J., S. Kubis, S. Bimanadham, and P. Jarvis. 2007. Functional similarity between the chloroplast translocon component, Tic40 and the human co-chaperone, Hsp70-interacting protein (Hip). *J. Biol. Chem.* 282: 21404–21414.

Berndt, K.D. 1996. Ultraviolet Circular Dichroism Spectroscopy and Application to Biological molecules. In: *Optical Spectroscopy and Dynamics of Biomolecular Structures.*

Bölter, B., T. May, and J. Soll. 1998. A protein import receptor in pea chloroplasts, Toc86, is only a proteolytic fragment of a larger polypeptide. *FEBS Lett*. 441: 59-62.

Brock, I.W., L. Hazell, D. Michl, V.S. Nielsen, B.L. Moller, R.G. Herrmann, R.B. Klosgen, and C. Robinson. 1993. Precursors of one integral and five lumenal thylakoid proteins are imported by isolated pea and barley thylakoids: Optimisation of *in vitro* assays. *Plant Mol. Biol.* 23: 717-725.

Bruce, B.D. 2001. The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim. Biophys. Acta.* 1541: 2-21.

Caliebe, A., R. Grimm, G. Kaiser, J. Lübeck, J. Soll, and L. Heins. 1997. The chloroplastic protein import machinery contains a Rieske-type iron–sulfur cluster and a mononuclear iron-binding protein. *EMBO J.* 16: 7342–7350.

Chen, K., X. Chen, and D.J. Schnell. 2000. Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. *Plant Physiol*. 122: 813-822.

Chen, D., and D.J. Schnell. 1997. Insertion of the 34-kDa chloroplast protein import component, IAP34, into the chloroplast outer membrane is dependent on its intrinsic GTP-binding capacity. *J. Biol. Chem.* 272: 6614-6620.

Chou, M.L., L.M. Fitzpatrick, S.L. Tu, G. Budziszewski, S. Potter-Lewis, M. Akita, J.Z. Levin, K. Keegstra, and H.M. Li. 2003. Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. *EMBO J.* 22: 2970-2980.

Chou, M.L., C.C Chu, L.J. Chen, M. Akita, and H.M. Li. 2006. Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. *J. Cell Biol.* 175: 893-900.

Constan, D., R. Patel, K. Keegstra, and P. Jarvis. 2004. An outer envelope membrane component of the plastid protein import apparatus plays an essential role in *Arabidopsis*. *Plant J.* 38: 93-106.

Dosztányi, Z., V. Csizmók, P. Tompa, and I. Simon. 2005a. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21: 3433-3434.

Dosztányi, Z., V. Csizmók, P. Tompa, and I. Simon. 2005b. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* 347: 827-839.

Dyall, S.D., M.T. Brown, and P.J. Johnson. 2004. Ancient invasions: From endosymbionts to organelles. *Science* 304: 253-257.

Dyson, H.J., and P.E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol*. 6: 197-208.

Fink, A.L. 2005. Natively unfolded proteins. *Curr. Opin. Struct. Biol*. 15: 35-41.

Finnie, C., J. Borch and D.B. Collinge. 1999. 14-3-3 proteins: eukaryotic regulatory proteins with many functions. *Plant Molec. Biol.* 40: 545-554.

Fulgosi, H., H. Lepeduŝ, V. Cesar, and N. Ljubeŝić. 2005. Differential accumulation of plastid preprotein translocon components during spruce (*Picea abies* L. Karst.) needle development. *Biol. Chem.* 386: 777-783.

Fulgosi, H., and J. Soll. 2001. The chloroplast protein import receptors Toc34 and Toc159 are phosphorylated by distinct protein kinases. *J. Biol. Chem*. 277: 8934-8940.

Fuxreiter, M., S. István, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Molec. Biol.* 338:1015-1026.

Gentle, I.E., L. Burri, and T. Lithgow. 2005. Molecular architecture and function of the Omp85 family of proteins. *Mol. Microbiol.* 58: 1216-1225.

Graceffa, P., A. Jancsó, and K. Mabuchi. 1992. Modifiction of acidic residues normalizes sodium dodecyl sulfate – polyacrylamide gel electrophoresis of caldesmon and other proteins that migrate anomalously. *Arch. Biochem. Biophys*. 297: 46-51.

Gutensohn, M., B. Schulz, P. Nicolay, and U.I. Flügge. 2000. Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus. *Plant J.* 23: 771-783.

Heins, L., A. Mehrle, R. Hemmler, R. Wagner, M. Kuchler, F. Hormann, D. Sveshnikov, and J. Soll. 2002. The preprotein conducting channel at the inner envelope membrane of plastids. *EMBO J.* 21: 2616–2625.

Hernández Torres, J., M.A.A. Maldonado, and J. Chomilier. 2007. Tandem duplications of a degenerated GTP-binding domain at the origin of GTPase receptors Toc159 and thylakoidal SRP. *Biochem. Biophys. Res. Commun.* 364: 325-331.

Hiltbrunner, A., J. Bauer, P.A. Vidi, S. Infanger, P. Weibel, M. Hohwy, and F. Kessler. 2001. Targeting of an abundant cytosolic form of the protein import receptor atToc159 to the outer chloroplast membrane. *J. Cell Biol.* 154: 309-316.

Hiltbrunner, A., K. Grünig, M. Alvarez-Huerta, S. Infanger, J. Bauer, and F. Kessler. 2004. AtToc90, a new GTP-binding component of the Arabidopsis chloroplast protein import machinery. *Plant Molec. Biol*. 54: 427-440.

Hinnah, S.C., R. Wagner, N. Sveshnikova, R. Harrer, and J. Soll. 2002. The chloroplast protein import channel Toc75: pore properties and interaction with transit peptides. *Biophys. J.* 83: 899-911.

Hirohashi, T., T. Hase, and M. Nakai. 2001. Maize non-photosynthetic ferredoxin precursor is mis-sorted to the intermembrane space of chloroplasts in the presence of light. *Plant Physiol.* 125: 2154-2163.

Hirsch, S., E. Muckel, F. Heemeyer, G. Von Heijne, and J. Soll. 1994. A receptor component of the chloroplast protein translocation machinery. *Science* 266: 1989-1992.

Hoare, D.G. and D.E. Koshland. 1967. A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* 242: 2447-2453.

Hofmann, N.R. and S.M. Theg. 2005. Toc64 is not required for import of proteins into chloroplasts in the moss *Physomitrella patens*. *Plant J.* 43: 675-687.

Hormann, F., M. Kuchler, D. Sveshnikov, U. Oppermann, Y. Li, and J. Soll. 2004. Tic32, an essential component in chloroplast biogenesis. *J. Biol. Chem.* 279: 34756–34762.

Inaba, T., M. Alvarez-Huerta, M. Li, J. Bauer, C. Ewers, F. Kessler and D.J. Schnell. 2005. Arabidopsis Tic110 is essential for the assembly and function of the protein import machinery of plastids. *Plant Cell.* 17: 1482-1496.

Inaba, T., M. Li, M. Alvarez-Huerta, F. Kessler, and D.J. Schnell. 2003. atToc110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *J. Biol. Chem.* 278: 38617-38627.

Inaba, T., and D.J. Schnell. 2008. Protein trafficking to plastids: one theme, many variations. *Biochem. J.* 413: 15-28.

Inoue, K., A.J. Baldwin, R.L. Shipman, K. Matsui, S.M. Theg, and M. Ohme-Takagi. 2005. Complete maturation of the plastid protein translocation channel requires a type I signal peptidase. *J. Cell Biol.* 171: 425-430.

Ivanova, Y., M.D. Smith, K. Chen, and D.J. Schnell. 2004. Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. *Mol. Biol. Cell.* 15: 3379-3392.

Jackson-Constan, D., and K. Keegstra. 2001. Arabidopsis genes encoding components of the chloroplastic protein import apparatus. *Plant Physiol*. 125: 1567-1576.

Jarvis, P., L.J. Chen, H.M. Li, C.A. Peto, C. Fankhauser, and J. Chory. 1998. An Arabidopsis mutant defective in the plastid general protein import apparatus. *Science.* 282: 100-103.

Jelic, M., N. Sveshnikova, M. Motzkus, P. Horth, J. Soll, and E. Schleiff. 2002. The chloroplast import receptor Toc34 functions as a preprotein-regulated GTPase. *Biol. Chem.* 383: 1875-1883.

Jelic, M., J. Soll, and E. Schleiff. 2003. Two Toc34 homologues with different properties. *Biochemistry.* 42: 5906-5916.

Kessler, F., G. Blobel, H.A. Patel, and D.J. Schnell. 1994. Identification of two GTP-binding proteins in the chloroplast protein import machinery. *Science.* 266: 1035-1039.

Kikuchi, S., T. Hirohashi, and M. Nakai. 2006. Characterization of the preprotein translocon at the outer envelope membrane of chloroplasts by Blue Native PAGE. *Plant Cell Physiol.* 47: 363-371.

Koenig, P., M. Oreb, A. Höfle, S. Kaltofen, K. Rippe, I. Sinning, E. Schleiff, and I. Tews. 2008. The GTPase cycle of the chloroplast import receptors Toc33/Toc34: Implications from monomeric and dimeric structures. *Structure.* 16: 585-596.

Kouranov, A., X. Chen, B. Fuks, and D.J. Schnell. 1998. Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. *J. Cell Biol.* 143: 991-1002.

Kouranov, A., and D.J. Schnell. 1997. Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. *J. Cell Biol.* 139: 1677-1685.

Kovacheva, S., J. Bedard, R. Patel, P. Dudley, D. Twell, G. Rios, C. Koncz, and P. Jarvis. 2005. *In vivo* studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. *Plant J.* 41: 412–428.

Kubis, S., A. Baldwin, R. Patel, A. Razzaq, P. Dupree, K. Lilley, J. Kurth, D. Leister, and P. Jarvis. 2003. The Arabidopsis *ppi1* mutant is specifically defective in the expression, chloroplast import, and accumulation of photosynthetic proteins. *Plant Cell.* 15: 1859-1871.

Kubis, S., R. Patel, J. Combe, J. Bédard, S. Kovacheva, K. Lilley, A. Biehl, D. Leister, G. Ríos, C. Koncz, and P. Jarvis. 2004. Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. *Plant Cell.* 16: 2059-2077.

Lakowicz, J.R. 2006. Principles of Fluorescence Spectroscopy Third Edition. NY: Plenum Publishers. 529-573.

Lee, K.H., S. J. Kim, Y.J. Lee, J.B. Jin, and I. Hwang. 2003. The M domain of atToc159 plays an essential role in the import of proteins into chloroplasts and chloroplast biogenesis. *J. Biol. Chem.* 278: 36794-36805.

Lee, K.H, D.H. Kim, S.W. Lee, Z.H. Kim, and I. Hwang. 2002. *In vivo* import experiments in protoplasts reveal the importance of the overall context but not specific amino acid residues of the transit peptide during import into chloroplasts. *Mol. Cells.* 14: 388-397.

Lee, D.W., S. Lee, G. Lee, K.H. Lee, S. Kim, G.W. Cheong, and I. Hwang. 2006. Functional characterization of sequence motifs in the transit peptide of Arabidopsis small subunit of Rubisco. *Plant Physiol.* 140: 466-483.

Ma, Y., A. Kouranov, S. LaSala, and D.J. Schnell. 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. *J. Cell Biol.* 134: 1-13.

Marcos-Lousa, C., D. P. Sideris, and K. Tokatlidis. 2006. Translocation of mitochondrial inner-membrane proteins: conformation matters. *Trends Biochem. Sci*. 31 :259-267.

Martin, W., and R.G. Hermann. 1998. Gene transfer from organelles to the nucleus: how much, what happens and why? *Plant Physiol.*188: 9-17.

Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A*. 99: 12246-12251.

Mathews, C.K. and K.E. Van Holde. 1996. Biochemistry 2nd Edition. San Francisco, CA: Benjamin-Cummings Publishing. 204-210.

May, T., and J. Soll. 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. *Plant Cell.* 12: 53-64.

Miras, S., D. Salvi, L. Piette, D. Seigneurin-Berny, D. Grunwald, C. Reinbothe, J. Joyard, S. Reinbothe, and N. Rolland. 2007. Toc159- and Toc75-independent import of a transit sequence-less precursor into the inner envelope of chloroplasts. *J. Biol. Chem.* 282: 29482-29492.

Nada, A., and J. Soll. 2004. Inner envelope protein 32 is imported into chloroplasts by a novel pathway. *J. Cell. Sci.* 117: 3975-3982.

Nanjo, Y., H. Oka, N. Ikarashi, K. Kaneko, A. Kitajima, T. Mitsui, F.J. Munoz, M. Rodriguez-Lopez, E. Boraja-Fernandez, and J. Pozueta-Romero. 2006. Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-Golgi to the chloroplast through the secretory pathway. *Plant Cell.* 18: 2582-2592.

Nielsen, E., M. Akita, J. Davila-Aponte, and K. Keegstra. 1997. Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO J.* 16: 935-946.

Perry, S.E., and K. Keegstra. 1994. Envelope membrane proteins that interact with chloroplastic precursor proteins. *Plant Cell.* 6: 93-105.

Pfanner, N. 2000. Protein sorting: Recognizing mitochondrial presequences. *Current Biol.* 10: R412-R415.

Prilusky, J., C.E. Felder, T. Zeev-Ben-Mordehai, E.H. Rydberg, O. Man, J.S. Beckmann, I. Silman, and J.L Sussman. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. 21: 3435-3438.

Puntervoll, P., R. Linding, C. Gemünd, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D.M.A. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferrè, V. Maselli, A. *Via*, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Küster, M. Helmer-Citterich, W.N. Hunter, R. Aasland, and T.J. Gibson. 2003. ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.,* 31: 3625-3630.

Qbadou, S., R. Tien, J. Soll, and E. Schleiff. 2003. Membrane insertion of the chloroplast outer envelope protein, Toc34: Constrains for insertion and topology. *J. Cell Sci.* 116: 837-846.

Raven, J.A., and J.F. Allen. 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol.* 4: 209.

Receveur-Bréchot, V., J.M. Bourhis, V.N. Uversky, B. Canard, and S. Longhi. 2006. Assessing protein disorder and induced folding. *Proteins.* 62: 24-45.

Reddick, L.E., M.D. Vaughn, S.J. Wright, I.M. Campbell, and B.D. Bruce. 2007. *In vitro* comparative kinetic analysis of the chloroplast Toc GTPases. *J. Biol. Chem.* 282: 11410-11426.

Reumann, S., K. Inoue, and K. Keegstra. 2005. Evolution of the general protein import pathway of plastids. *Mol. Membrane Biol.* 22: 73-86.

Roccatano, D., G. Colombo, M. Fioroni, and A.E. Mark. 2002. Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: A molecular dynamics study. *Proc. Natl. Acad. Sci. U.S.A*. 99: 12179-12184.

Schleiff, E., and R.B. Klösgen. 2001. Without a little help from 'my' friends: direct insertion of proteins into chloroplast membranes? *Biochim. Biophys. Acta.* 1541: 22–33.

Schleiff, E., M. Jelic, and J. Soll. 2003. A GTP-driven motor moves proteins across the outer envelope of chloroplasts. *Proc. Natl. Acad. Sci.U.S.A.* 100: 4604-4609.

Schleiff, E., J. Soll, W. Kühlbrandt, and R. Harrer. 2003. Characterization of the translocon of the outer envelope of chloroplasts. *J. Cell Biol.* 160: 541-551.

Schnell, D.J., F. Kessler, and G. Blobel. 1994. Isolation of components of the chloroplast protein import machinery. *Science.* 266: 1007-1012.

Schulz, A., J. Knoetzel, H.V. Scheller, and A. Mant. 2004. Uptake of a fluorescent dye as a swift and simple indicator of organelle intactness: import-competent chloroplasts from soil-grown Arabidopsis. *J. Histochem. Cytochem.* 52: 701-704.

Seedorf, M., K. Waegemann, and J. Soll. 1995. A constituent of the chloroplast import complex represents a new type of GTP-binding protein. *Plant J.* 7: 401-411.

Smith, M.D. 2006. Protein import into chloroplasts: an ever-evolving story. *Can. J. Bot.* 84: 531-541.

Smith, M.D., L. Fitzpatrick, K. Keegstra, and D.J. Schnell. 2002a. *In vitro* analysis of chloroplast protein import. In: *Current Protocols in Cell Biology*. Eds. J. S. Bonifacino, M. Dasso, J. B. Harford, J. Lippincott-Schwartz, and K.M. Yamada. New York: John Wiley and Sons Inc. 11.16.1-11.16.21.

Smith, M.D., A. Hiltbrunner, F. Kessler, and D.J. Schnell. 2002b. The targeting of the atToc159 preprotein receptor to the chloroplast outer membrane is mediated by its GTPase domain and is regulated by GTP. *J. Cell Biol.* 159: 833-843.

Smith, M.D., C.M. Rounds, F. Wang, K. Chen, M. Afitlhile, and D.J. Schnell. 2004. AtToc159 is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. *J. Cell Biol.* 165: 323-334.

Smith, M.D., and Schnell, D.J. Chloroplast Protein Targeting: Multiple Pathways for a Complex Organelle. In: *Protein Movement Across Membranes*. Ed. J. Eichler. 2004. Georgetown, TX: Landes Biosciences. 1-18.

Sohrt K., and J. Soll. 2000. Toc64, a new component of the protein translocon of chloroplasts. *J. Cell. Biol.* 148: 1213-1221.

Stengel, A., P. Benz, M. Balsera, J. Soll, and B. Bolter. 2008. TIC62 redox-regulated translocon composition and dynamics. *J. Biol. Chem.* 283: 6656-6667.

Teng, Y.S., Y.S. Su, L.J. Chen, Y.J. Lee, I. Hwang, and H.M. Li. 2006. Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell.* 18: 2247-2257.

Thompson, W.W., and J.M. Whatley. 1980. Development of non-green plastids. *Annu. Rev. Plant Physiol.* 31: 375-394

Tompa, P. 2002. Intrinscially unstructured proteins. *Trends Biochem. Sci.* 27: 527-533.

Tompa, P. 2003. Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays.* 25: 847-855.

Tompa, P., and M. Fuxreiter. 2007. Fuzzy complexes: polymorphisms and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33: 2-8.

Tranel, P., J. Froehlich, A. Goyal, K. Keegstra. 1995. A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane *via* a novel pathway. *EMBO J.* 14: 2436–2446.

Tranel, P., and K. Keegstra. 1996. A novel, bipartite transit peptide targets OEP75 to the outer membrane of the chloroplastic envelope. *Plant Cell.* 8: 2093–2104.

Tsai, L.Y., S.L. Tu, and H.M. Li. 1999. Insertion of atToc34 into the chloroplastic outer membrane is assisted by at least two proteinaceous components in the import system. *J. Biol. Chem.* 274: 18735-18740.

Uversky, V. 2002a. What does it mean to be natively unfolded? *Eur. J. Biochem*. 269: 2-12.

Uversky, V. 2002b. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* 11: 739-756.

Villarejo, A., S. Buren, S. Larsson, A. Dejardin, M. Monne, C. Rudhe, J. Karlsson, S. Jansson, P. Lerouge, N. Rolland, G. von Heijne, M. Grebe, L. Bako, and G. Samuelsson. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat. Cell Biol.* 7: 1224-1231.

Voigt, A., M. Jakob, R.B. Klösgen, and M. Gutensohn. 2005. At least two Toc34 protein import receptors with different specificities are also present in spinach chloroplasts. *FEBS Letters* 579: 1343-1349.

von Heijne, G., J. Steppuhn, and R.G. Herrmann. 1989. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* 180: 535-545.

Wallas, T.R., M.D. Smith, S. Sanchez-Nieto, and D.J.Schnell. 2003. The roles of Toc34 and Toc75 in targeting the Toc159 preprotein receptor to chloroplasts. *J. Biol. Chem.* 278: 44289-44297.

Whitmore, L., and B.A. Wallace. 2004. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32: W668-673.

Williams, A., and V. Frasca. 1998. Ion-Exchange Chromatography. In: *Current protocols in molecular biology*. John Wiley and Sons Inc. 10.10.1-10.10.30.

Wise, R.R. 2006. The Diversity of Plastid Form and Function. In: *The Structure and Function of Plastids*. Eds. R.R. Wise, and J.K. Hoober. Dordrecht: Springer. 3-26

Yan, X., S. Khan, T. Hase, M.J. Emes, and C.G. Bowsher. 2006. Differential uptake of photosynthetic and non-photosynthetic proteins by pea root plastids. *FEBS Lett*. 580: 6509-6512.

Appendices

APPENDIX 1: *In vitro* targeting assay data

Binding Efficiency Summary

	WT			*ppi1*			*ppi3*		
	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]
atToc159	6.57	1.76	0.65	1.71	0.73	0.30	2.44	1.11	0.45
159GM	6.36	1.51	0.51	4.83	1.24	0.51	3.12	1.66	0.68
132A159GM	7.07	1.35	0.38	4.01	1.84	0.75	4.19	0.50	0.20
atToc132	6.05	2.11	0.95	2.43	1.16	0.47	4.19	2.09	1.05
132GM	10.97	3.87	1.26	8.24	1.41	0.58	4.53	0.66	0.27
159A132GM	2.98	0.69	0.22	0.78	0.17	0.07	2.93	1.04	0.42

[a] % *in vitro* translation product added to the targeting reaction; SD = standard deviation; SEM = standard error of the mean

Inserted:Bound Summary

	WT			*ppi1*			*ppi3*		
	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]
atToc159	0.50	0.27	0.11	0.27	0.12	0.05	0.34	0.08	0.03
159GM	0.27	0.04	0.02	0.03	0.05	0.02	0.10	0.07	0.03
132A159GM	0.46	0.13	0.05	0.39	0.27	0.11	0.30	0.30	0.11
atToc132	0.22	0.10	0.04	0.15	0.07	0.04	0.10	0.04	0.02
132GM	0.16	0.07	0.02	0.13	0.04	0.02	0.15	0.12	0.05
159A132GM	0.16	0.04	0.01	0.15	0.08	0.03	0.19	0.07	0.03

[a] ratio calculated based on corrected radioactivity CNT; SD = standard deviation; SEM = standard error of the mean

Insertion Efficiency Summary

	WT			*ppi1*			*ppi3*		
	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]	Mean[a]	SD	SEM[a]
atToc159	3.50	0.89	0.26	0.53	0.20	0.08	0.82	0.34	0.12
159GM	2.81	1.34	0.47	0.11	0.17	0.07	0.18	0.21	0.09
132A159GM	3.37	0.92	0.33	1.61	1.16	0.47	1.36	1.30	0.53
atToc132	1.28	0.78	0.40	0.36	0.21	0.09	0.45	0.33	0.13
132GM	1.90	1.14	0.40	1.04	0.33	0.13	0.69	0.16	0.07
159A132GM	0.46	0.13	0.05	0.11	0.03	0.01	0.44	0.12	0.05

[a] % *in vitro* translation product added to the targeting reaction; SD = standard deviation; SEM = standard error of the mean

APPENDIX 2: Statistical analyses

One-way Analysis of Variance Summary

Parameter	Constructs	Chloroplast type	d.f. (between groups)	d.f. (within groups)	F	P
Binding Efficiency (% IVTP)[a]	atToc159-derived	WT	2	19	0.569	0.576
	atToc132-derived	WT	2	18	21.27	< 0.001*
	atToc159-derived	*ppi1*	2	15	8.59	0.003*
	All Constructs	*ppi3*	5	28	3.33	0.018*
Inserted:Bound	All Constructs	WT, *ppi1*, *ppi3*	8	46	1.02	0.433
	atToc159-derived	WT	2	13	1.85	0.196
Insertion Efficiency (% IVTP)[a]	atToc159-derived	WT	2	21	0.914	0.416

[a] IVTP = *in vitro* translation product added to targeting reaction; *significant differences at $P \leq 0.05$

Two-way Student's t-test (unequal variance) results

WT binding	*atToc132*	*132GM*
Mean	6.054	10.974
Variance	4.466	12.700
Observations	5.000	8.000
Pooled Variance	9.706	
Hypothesized Mean Difference	0.000	
df	11.000	
t Stat	-2.770	
P(T<=t) one-tail	0.009	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.018	
t Critical two-tail	2.201	

WT binding	*atToc132*	*159A132GM*
Mean	6.054	2.979
Variance	4.466	0.400
Observations	5.000	8.000
Pooled Variance	1.879	
Hypothesized Mean Difference	0.000	
df	11.000	
t Stat	3.936	
P(T<=t) one-tail	0.001	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.002	
t Critical two-tail	2.201	

WT inserted:bound	*atToc132-derived*	*atToc159-derived*
Mean	0.171	0.326
Variance	0.005	0.046
Observations	21.000	16.000
Hypothesized Mean Difference	0.000	
df	17.000	
t Stat	-2.790	
P(T<=t) one-tail	0.006	
t Critical one-tail	1.740	
P(T<=t) two-tail	0.013	
t Critical two-tail	2.110	

WT insertion efficiency (%IVTP)	*atToc159-derived*	*atToc132-derived*
Mean	3.169	1.207
Variance	1.225	0.977
Observations	24.000	22.000
Hypothesized Mean Difference	0.000	
df	44.000	
t Stat	6.351	
P(T<=t) one-tail	0.000	
t Critical one-tail	1.680	
P(T<=t) two-tail	0.000	
t Critical two-tail	2.015	

ppi1 binding	*132A159GM*	*atToc159*
Mean	4.010	1.713
Variance	3.403	0.533
Observations	6.000	6.000
Hypothesized Mean Difference	0.000	
df	7.000	
t Stat	2.835	
P(T<=t) one-tail	0.013	
t Critical one-tail	1.895	
P(T<=t) two-tail	0.025	
t Critical two-tail	2.365	

ppi1 binding	*atToc159*	*159GM*
Mean	1.713	4.828
Variance	0.533	1.526
Observations	6.000	6.000
Hypothesized Mean Difference	0.000	
df	8.000	
t Stat	-5.317	
P(T<=t) one-tail	0.000	
t Critical one-tail	1.860	
P(T<=t) two-tail	0.001	
t Critical two-tail	2.306	

ppi1 inserted:bound	*atToc159*	*132A159GM*
Mean	0.271	0.390
Variance	0.015	0.077
Observations	6.000	6.000
Hypothesized Mean Difference	0.000	
df	7.000	
t Stat	-0.959	
P(T<=t) one-tail	0.185	
t Critical one-tail	1.895	
P(T<=t) two-tail	0.370	
t Critical two-tail	2.365	

ppi1 inserted:bound	*atToc159*	*159GM*
Mean	1.713	4.828
Variance	0.533	1.526
Observations	6.000	6.000
Hypothesized Mean Difference	0.000	
df	8.000	
t Stat	-5.317	
P(T<=t) one-tail	0.000	
t Critical one-tail	1.860	
P(T<=t) two-tail	0.001	
t Critical two-tail	2.306	

APPENDIX 3: Primers used for PCR amplification of the A-domain of atToc159 and atToc132.

	Primer Set	Purpose of Primer Set		Sequence
atToc159A Fragment 1	1	Incorporate codons at 5' end for 4 His residues and 3' *Kpn I* restriction site for subcloning purposes	S	5' AC**CACCACCACCAC**GACTCAAAGTCGGTTACT 3'
			AS	5' CCTCTTTGGTACCATTGTCA 3'
	2	Incorporate remainder of 5' His tag and *Nhe I* restriction site for subcloning purposes	S	5' GGGGCTAGC**CACCACCACCACCACCAC**GACTCA 3'
			AS	5' CCTCTTTGGTACCATTGTCA 3'
atToc159A Fragment 2	3	Incorporate 5' *Kpn I* and 3' *Sal I* restriction sites for subcloning purposes	S	5' TGACAATGGTACCAAAGAGG 3'
			AS	5' CCCGTCGACTCATGTTATGGTAAAATTGCC 3'
atToc132A	4	Incorporate codons at 5' end for 4 His residues	S	5' AC**CACCACCACCAC**GGAGATGGGACTGAGTTT 3'
			AS	5' CCCGAGCTCTCAAAGACCTGCTGGACGAGCAG 3'
	5	Incorporate remainder of 5' His tag and 3' *Sac I* restriction site for subcloning purposes	S	5' GGGGCTAGC**CACCACCACCACCACCAC**GGAGAT 3'
			AS	5' CCCGAGCTCTCAAAGACCTGCTGGACGAGCAG 3'

Note: Sequences corresponding to codons for N-terminal Histidine residues are shown in bold; restriction enzyme sites are underlined; S = sense primer; AS = antisense primer