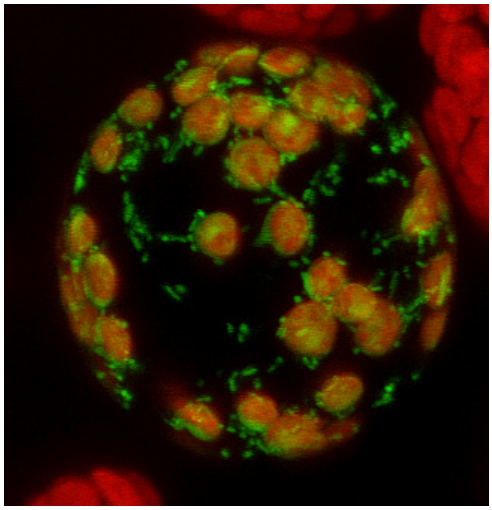
UNIVERSITY OF TROMSØ UiT

Faculty of Biosciences, Fisheries and Economics
Department of Arctic and Marine Biology

Plant regulatory networks:

RNA binding proteins as mediators of communication between DNA containing compartments by dual targeting



Janina Fuss

A dissertation for the degree of
Philosophiae Doctor

May 2012

Plant regulatory networks:

RNA binding proteins as mediators of communication between DNA containing compartments by dual targeting

Janina Fuss

May 2012

Contents

Acknowledgement

This study was done in the Department for Arctic and Marine Biology at the Faculty for Biosciences, Fisheries and Economics in the Molecular Environments Group (formerly known as the Microbiology and Plant physiology) under supervision of Prof. Kirsten Krause and Prof. Karsten Fischer and financed by the Norwegian research council.

First of all, I would like to thank Kirsten Krause for her supervision, her dedication and her enthusiasm concerning new ideas (hers and mine) and her support in finding new angles and new partners for collaboration. Our long discussions opened the way for a huge variety of new insights, which brought not only my project, but also me forward.

I also thank my second supervisor Karsten Fischer for critical comments and bringing us back to reality, whenever the ideas and speculations turned too absurd.

I also want to thank my collaboration partners in Köln (Prof. Flügge and Dr. Gigolashvili) and Freiburg (Prof. Stefan Rensing, Oliver Liegmann and Jan Mitschke), who gave me the possibility to learn new techniques, use different plant systems and who developed the prediction tools. I want to thank Rainer Schwacke not only for helping with bioinformatics questions and problems, but also for influencing my weekend activities by planning extensive hikes.

A really huge thank you goes to the technicians and green house staff Rigmor, Coby, Leidulf and Alena, who helped a lot with the plant work and with the yeast work, including fighting bugs of all kinds who tried to disturb growth.

A thank you of the same size goes to the students (Stina, Anja, Ina, Kathrin and, a long time ago, Steingrim and Steffi) in the lab, who helped with cloning, screening, preparation of solutions and media or worked on their own projects and really did a great job.

But all that work wouldn't have been fun without everyone already mentioned and all the others in the lab, who "shortened" incubation times, helped with technical and organizational problems and everything else that came up, including the really "depressing" Friday event vinlotteri.

Last but not least, I want to thank all people who helped with corrections and discussions, suffered through even worse moods than usual and still talk to me on a voluntary basis, especially my friends, family and the person, who helped most and probably also suffered most, Ulli.

Abstract

The endosymbiosis-derived organelles within a plant cell, plastids and mitochondria, have to be equipped with a certain set of proteins to be fully functional. This set of proteins is encoded by different genomes: the organellar genomes and the nuclear genomes. This setup poses some interesting challenges for the regulation of gene expression and protein transport. On the one hand, the targeting signals that transport proteins to the organelles have to be highly specific and on the other hand, the communication between the DNA containing compartments to coordinate their gene expression has to be transmitted somehow, not only from the nucleus but also from the organelles back to the nucleus.

In this thesis, two prediction programs are introduced. One of them can predict dual targeted proteins to both organelles (ATP, **a**mbiguous **t**argeting **p**redictor) and the other one is species- specific for *Physcomitrella patens* (GTP_Pp; **g**reen **t**argeting **p**redictor – *P. patens*-specific). The first predictor can help to gain a more complete picture of the proteins potentially present in the organelles. With the help of that predictor, we predicted that the amount of proteins with dual targeting signals is higher than anticipated and that we to date know only a minor part of actually dual targeted proteins. The second predictor can help to answer the question on the evolutionary consistency of targeting signals within the plant kingdom and the importance of having species-specific approaches in analyzing protein targeting. We actually observed a surprisingly big difference in composition and recognition of mitochondrial and dual-targeting protein signals, which led to the conclusion, that species-specific approaches always should be considered as the optimal option for both, *in silico* and *in vivo* experiments.

The second part of this thesis focuses the mechanisms of communication between nucleus and the organelles, especially the plastid possibly mediated by dual targeting. We chose several plastid RNA-/DNA-binding proteins to analyze their sub-plastidic localization and their potential additional nuclear localization. Those candidates were AtWHY1 (*Arabidopsis thaliana* Whirly1), four members of the AtcpRNP (chloroplast ribonucleoprotein) family and AtEF-Tu (elongation factor thermo-unstable). The analyzed members of the AtcpRNP family reflect their described multiple functions within the plastid also in a multiple localization pattern. Furthermore, we were able to show interactions of different members of the AtcpRNP family by yeast-two-hybrid interaction assays. The localization pattern of AtEF-Tu was very similar to the one observed for the AtcpRNPs, which indicated, together with a confirmed localization within the transcriptionally active chromosome, a multiple function for AtEF-Tu. The sub-plastidic localization data suggest overlapping networks of activity for the proteins by observed co-localizations. This was also shown with respect to several marker proteins for plastid functions. For AtEF-Tu and the AtcpRNPs, we also showed experimentally that a second localization in the nucleus is possible for the mature protein, which makes them interesting candidates for a possible mediation of plastid signals to the nucleus next

to AtWHY1. For AtWHY1, we were able to show an effect of the DNAbinding domain on the known localization pattern which seemed to reflect an aberration in transport processes through the envelope. This offers a potential regulatory mechanism that needs to be explored in detail in the future.

List of publications and manuscripts

Paper I

Prediction of dual protein targeting to plant organelles

Mitschke J, Fuss J, Blum T, Hoglund A, Reski R, Kohlbacher O, Rensing SA. 2009.

New Phytol **183**(1): 224-235.

This publication is partially based on data that was included in my Diploma thesis.

Paper II

Whirly proteins as communicators between plant organelles and the nucleus?

Krause K, Herrmann U, Fuss J, Miao Y, Krupinska K. 2009.

Endocytobiosis Cell Res. **19**: 51-62.

Manuscript I

Moonlighting in plastids - translation elongation factor EF-Tu is a component of chloroplast transcriptionally active chromosomes.

Herrmann U, Fuss J, Krupinska K, Krause K. 2012.

Manuscript submitted to Current Genetics

Manuscript II

Co-Localization and interaction of ribonucleoproteins in chloroplasts of *Arabidopsis thaliana*.

Fuss J, Krause K. 2012

Manuscript in preparation

Manuscript III

Can *Arabidopsis thaliana* read the messages from *Physcomitrella patens* and *vice versa*? - An analysis on the conservation of targeting signals.

Fuss J, Liegmann O, Krause K, Rensing SA. 2012.

Manuscript submitted to New Phytologist

Manuscript IV "work in progress"

The DNA binding domain of a Whirly protein from *Arabidopsis thaliana* is engaged in protein translocation across the plastid envelope membrane.

Fuss J, Krause K. 2012.

Work in progress - manuscript in preparation.

I. Introduction

1.5 - 2 billion years ago, a prokaryotic or eukaryotic cell (Martin & Russell, 2003) took up a free living α-proteobacterium, and in the following million years, the symbiont turned into an endosymbiont with all consequences for both cells (Dyall *et al.*, 2004): the host cell gained the possibility to use oxygen for respiration, while the endosymbiont lived in a nutrition-saturated environment at the expense of providing energy for the host cell. The step towards becoming an organelle was fulfilled when the transfer of genomic material occurred from the symbiont towards the nucleus of the host cell and a mechanism for inheritance was established (Gross & Bhattacharya, 2009). This did not only bind the α-proteobacterium to the host forever, but also led to major organizational challenges for the host cell. The now nuclear-encoded and cytosol-translated proteins were still needed within the organelles we know today as mitochondria. This led to the development of a targeting/import system for proteins from the nucleus to the mitochondria, which had to be highly specific and efficient (Dolezal *et al.*, 2006). This import system was challenged then 1.2 – 1.5 billion years ago by a second uptake of a bacterium (reviewed by Dyall *et al.* (2004)). This second uptake of a cyanobacterium led to the origin of the green lineage (reviewed by Palmer (2003)). The resulting genomic reorganization made it necessary not only to develop new import mechanism for the new endosymbiont, the plastid, but also to refine the targeting mechanism of nuclear proteins towards the now two organelles (Macasev *et al.*, 2000; Bhattacharya *et al.*, 2007; Gross & Bhattacharya, 2009). As most processes were highly specialized, the specificity had to be given even though a subset of proteins was needed in both organelles like RNA-polymerases and other components of the transcriptional and translational machinery (e.g. Hedtke *et al.* (1997)). The main functions of both organelles in most seed plants, respiration in the mitochondria and photosynthesis in the plastids, are only one part of the specialization. The progressing compartmentalization of the cell led also to a complex network of metabolic processes taking place over several compartments (Neuhaus & Emes, 2000; Bowsher & Tobin, 2001; Padmasree *et al.*, 2002). An example for a plant-specific metabolic pathway that includes not only plastids but also mitochondria is the tetrapyrrol pathway. Its endproducts are amongst others chlorophyll and haem, which are mainly produced in the plastids, but the last two steps of haem synthesis happen in both organelles (Papenbrock & Grimm, 2001). Deciphering this kind of complex metabolic networks and their regulation is important because understanding plant signaling and production might help to use plants more efficiently. The main challenge for the regulation of the metabolic processes within a plant cell is the coordination of the three genomes, which are characteristic for plastid bearing organisms. Their tight regulation is not yet understood and the main questions to answer are:

How many and which proteins are present in which compartment at what time and why?

And:

How does the nucleus communicate with the organelles and vice versa*?*

Answering these questions is crucial to understand plant regulatory networks and the adaptions of plants to external signals.

PROTEIN TARGETING MECHANISMS

One of the most important steps towards answering the first question is the understanding of targeting mechanisms for proteins in general and the regulatory mechanism for proteins under different conditions. The understanding of those mechanisms is especially important as we find proteins that are localized to one compartment (single targeting) or more (dual targeting) in plant cells (Fig. 1) (Karniely & Pines, 2005).

Single targeting

To be able to understand reliably, which proteins occur in which compartment of the cell, several approaches are possible: localization experiments with tagged proteins, import experiments with isolated organelles (where possible) or whole compartment proteomics analyses by mass spectrometry to name the most prominent ones. Those approaches deliver information on a subset of proteins, but it is almost impossible to cover the complete range of proteins within the cell with those methods, even though mass spectrometry is a high throughput approach. The processes within the cell are extremely dynamic so that each mass spectrometry approach only depicts a snapshot of the situation within the organelle/compartment under these circumstances in this organism at that time. This provides very valuable knowledge on specific adaptions for specific treatments/processes, but, unfortunately, it has some flaws, too. First of all, the isolation of whole organelles and more, the isolation of compartments is prone to contaminations and to missing only loosely associated proteins (e.g. Majeran *et al.* (2012)). Furthermore, the isolation method actually might introduce a bias towards which proteins are isolated (Hu *et al.*, 2005), not to mention that usually a wide range of cell types with diverging functions and potentially diverging developmental statuses are isolated.

The other two methods are difficult to adapt to high throughput approaches and they do have some drawbacks, too: imports into isolated organelles are not always reflecting the actual situation *in vivo*, as within a cell, different receptors belonging to several import systems might be able to bind to the targeting signal, but some more efficiently than others (Cleary *et al.*, 2002). This difference in efficiency is not reflected in approaches with isolated organelles and an ambiguous picture might be the result (Krause *et al.*, 2005). This obstacle was overcome by the use of a mixture of isolated organelles (Rudhe *et al.*, 2002). The second problem are cytoplasmic factors that might facilitate or inhibit import into the compartment (Pfeiffer *et al.*, 2009). Those might not be present in isolated organelles or at least not in the physiological composition. Microscopy approaches with tagged proteins are used in *in vivo* systems and are thought to be possibly the most accurate system to show targeting (Millar *et al.*, 2006), but also there are some disadvantages: the attachment of a fusion protein might change the behavior of the protein *in vivo* either by changing the conformation of the protein or by masking a potential (second) targeting signal at the fusion site and so changing the complete functionality of the protein (Thomas & Maule, 2000). To avoid large tags, small tags and antibody labeling might be used, but this only works in fixed tissues and so the advantage of *in vivo* localization studies is abolished.

All methods have in common that they do not help to understand the underlying mechanisms. Separate approaches have been started in order to understand general characteristics of targeting mechanisms like what distinguishes mitochondrial from plastid targeting signals (Huang *et al.*, 2009), what is characteristic for the secretory pathway or what impact the C-terminus or internal characteristics have on the protein targeting (Dabney-Smith *et al.*, 1999; Bahaji *et al.*, 2011). The information on localizations gained from experimental methods in combination with information on mechanisms of targeting led to the development of targeting prediction programs like TargetP (Emanuelsson *et al.*, 2000), Predotar (Small *et al.*, 2004), MultiLoc (Hoglund *et al.*, 2006), SherLoc2 (Briesemeister *et al.*, 2009) or WolfPSORT (Horton *et al.*, 2007), to name just a few. All those programs use different kinds of approaches to predict the subcellular localization like neural networks, machine learning tools and/or literature and homology searches. Some of those predictors focus only on N-terminal signals, while other also predict internal signals like the nuclear localization signal (NLS) or C-terminal signals like those for peroxisomes. For the NLS and peroxisomal signals, a number of sequence motifs are known that encode translocation signals for the respective compartment, so the prediction is mainly based on similarity to known motifs and research aims at identifying additional motifs (Lingner *et al.*, 2011). In contrast to the sequence-based signals for the nucleus and the peroxisomes, N-terminal targeting signals are not conserved on primary sequence level, but rather on a secondary structure level (Bruce, 2001). The localizations, N-terminal targeting signals encode for, are plastids, mitochondria and the secretory pathway (Fig. 1). Especially plastid and mitochondrial sequences have been shown to be quite similar, even though there seem to be similarities to secretory proteins, too (von Heijne *et al.*, 1989; Huang *et al.*, 2009).

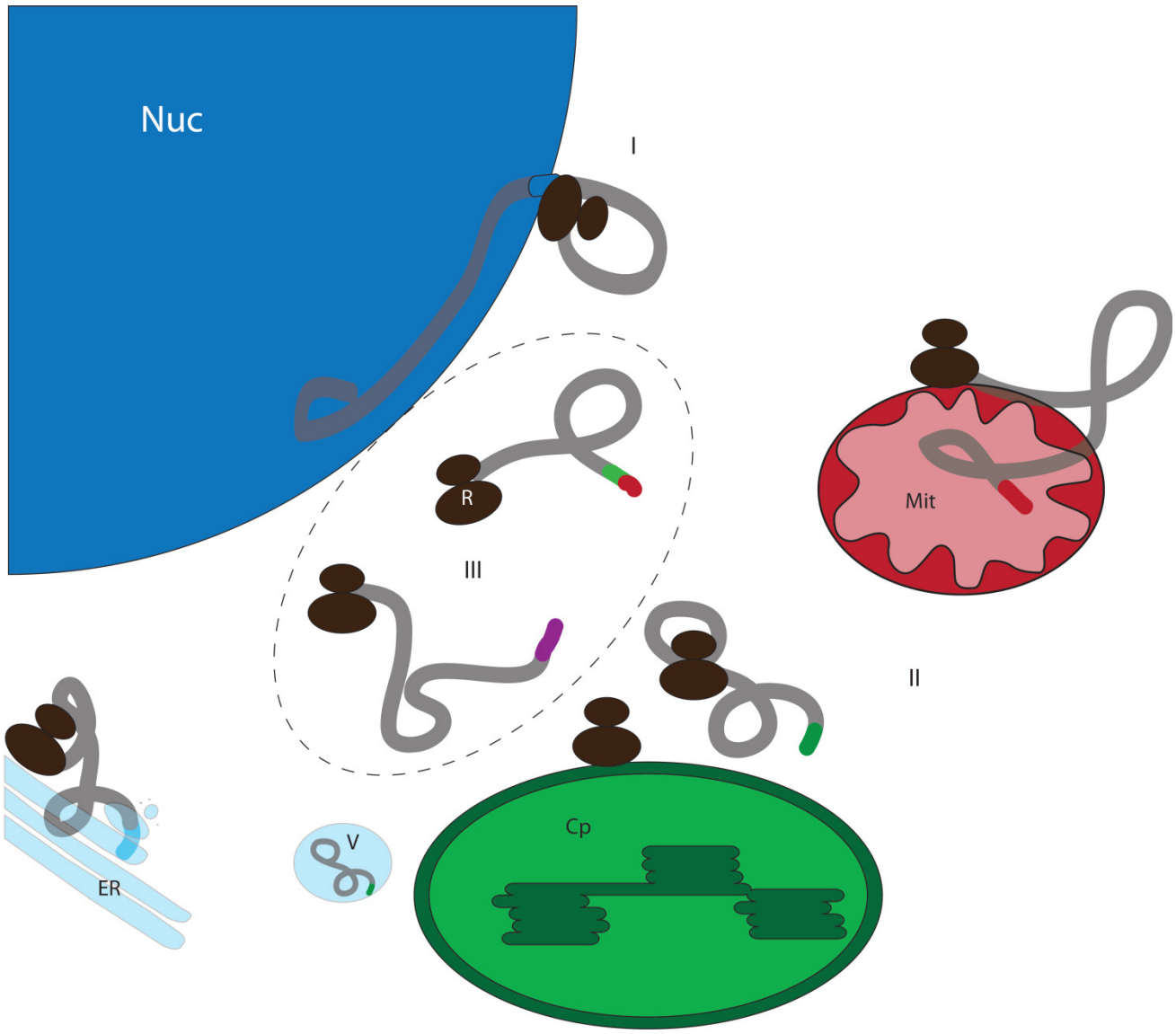


Figure 1: Different mechanisms of single or dual targeting are shown. Transcription takes place in the nucleus (Nuc), while translation of proteins (grey structures) takes place in the cytoplasm on either membrane associated or free ribosomes (R). Depicted here are N-terminal targeting signals: green for plastid (Cp), red for mitochondria (Mit) or light blue for secretory pathway (ER) (II) and an internal nuclear localization signal (dark blue) (I). Combinations of N-terminal targeting signal are given as twin signal to mitochondria and plastids (red and green) or as ambiguous targeting signal (purple)(III). It has been shown that some proteins reach the organelles via the secretory pathway, most likely via vesicles (V). This is exemplified for a plastid protein.

Dual targeting

It was only in the beginning of the 1990s that dual targeting as a possibility started to be described at all. Until then, the dogma "one protein, one compartment" was believed to be true. In the beginning it was only yeast mitochondrial proteins that were described to be imported into the plastids of plant cells (Huang *et al.*, 1990). In 1995, with the pea glutathione reductase, the first plant protein exhibiting dual targeting to plastids and mitochondria was described (Creissen *et al.*, 1995). Later on, the family of tRNA-sythetases was closer examined and one member was found to be dual targeted to the cytoplasm and the mitochondria (Mireau *et al.*, 1996). Step by step, it became clear that a large amount of those proteins are actually targeted to more than one compartment and that for this protein family, dual targeting is rather the rule than the exception (Duchene *et al.* (2005) and others). This realization came more or less hand in hand with an emerging number of newly sequenced plant genomes (Arabidopsis Genome, 2000; Goff *et al.*, 2002) whose annotations were and still are an ongoing process. This new information on multiple compartments for single proteins actually challenged the protein annotation to a large extent, as predictions were only possible for single targeted proteins. The few known examples at that time, that were described to be targeted to both, mitochondria and plastids, were analyzed quite detailed and a distinction between several ways of realizing dual targeting was made (Peeters & Small, 2001; Karniely & Pines, 2005; Millar *et al.*, 2006). Mechanisms mediating dual targeting are very diverse and regulation seems possible on many levels for N-terminal targeting signals. Even though the starting point for all processes is only one gene there are plenty of possibilities for organizing and regulating N-terminal dual targeting signals, roughly distinguished by the terms ambiguous targeting signals (Fig. 2a) or twin targeting signals (Fig. 2b-e).

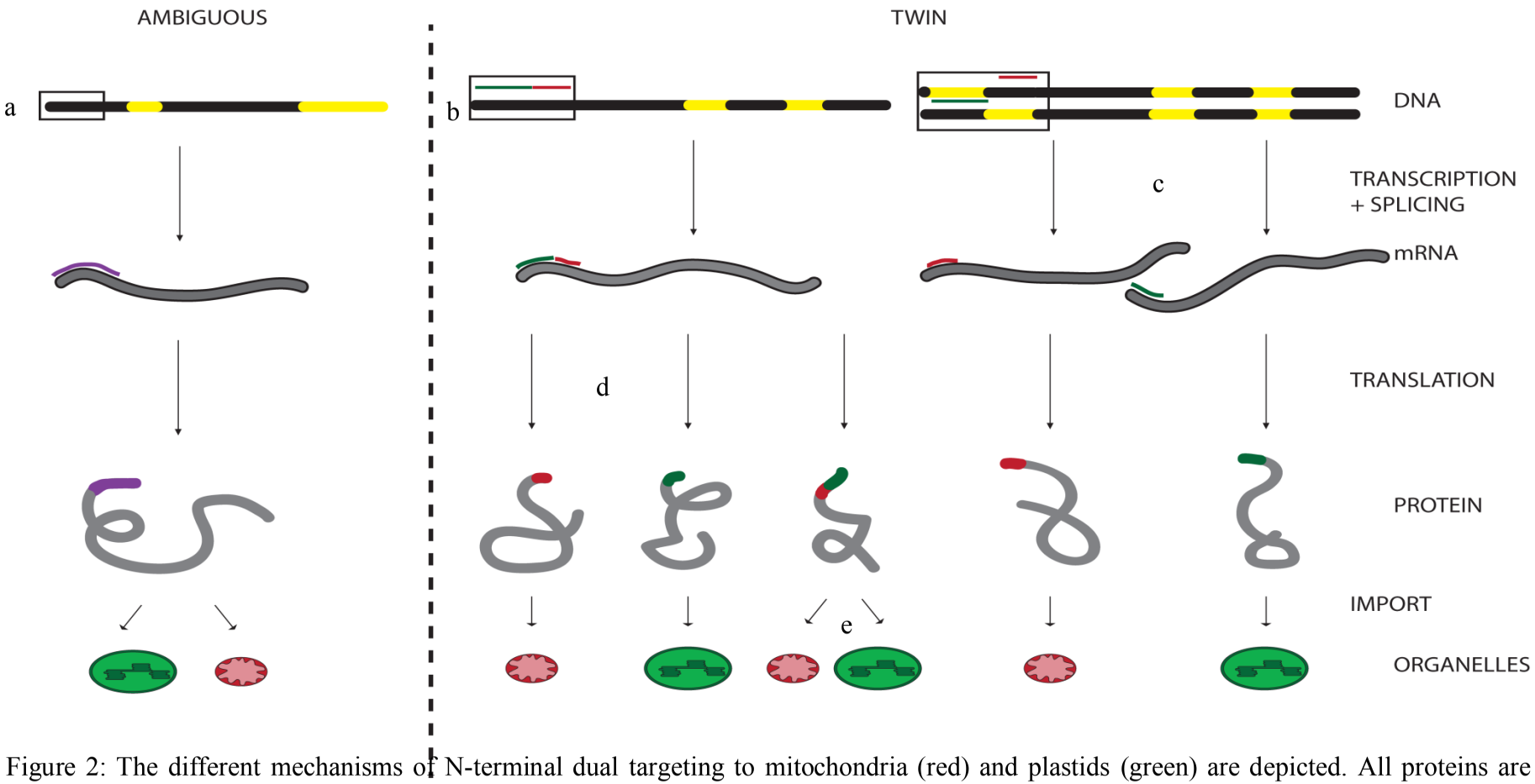


Figure 2: The different mechanisms of N-terminal dual targeting to mitochondria (red) and plastids (green) are depicted. All proteins are encoded by one gene. The DNA is represented in its intron (yellow) and exon (black) structure, while the black box shows in which region the targeting signal incl. 5' UTR (untranslated region) is encoded. For ambiguous targeting, only one mRNA (grey with black borders) is transcribed and one protein (grey) translated. The targeting signal is a mixture of characteristics (purple) and therefore recognized by both import machineries (a). Twin targeting is a lot more diverse. Targeting signals are arranged in a row (b) and two different transcription initiation sites might be used or one targeting signal might be spliced out so that two single targeted proteins originate from one gene (black box: whole targeting signal, green: plastid part, red: mitochondrial part) (c). When no splicing occurs, alternative start codons might lead to the translation of two single targeted proteins (d). When both targeting signals are translated, only one protein with two targeting signals in a row originates from the one gene (e).

Twin targeting signals

Twin targeting signals are actually two single targeting signals in a row, usually characterized by several in frame start-codons within the N-terminal part (Fig. 2b). The first time an actual import with a tandem arrayed targeting signal was demonstrated, an artificial twin construct was used. For this experiment, two single targeting sequences in a row were fused to a reporter gene. One directed the protein to mitochondria and one to plastids and reporter gene activity was shown in both compartments (de Castro Silva Filho *et al.*, 1996). The first actual twin targeting sequences were described for the Protoporphyrinogen oxidase of spinach (Watanabe *et al.*, 2001) and for the *A. thaliana* THI1 (Chabregas *et al.*, 2003). The regulatory possibilities, which targeting signal will be used, are again diverse. Multiple transcription initiation sites are in action (Obara *et al.*, 2002) as well as alternative splicing (Dinkins *et al.*, 2008) for regulation on mRNA-level (Fig. 2c). On translational level, multiple initiation sites are a regulatory option (Watanabe *et al.*, 2001; Chabregas *et al.*, 2003), some of which might not even be ATG-encoded. The decision on which starting point is used seems to be strongly influenced by the 5' UTR (untranslated region) (Christensen *et al.*, 2005) (Fig. 2d). All of those mechanisms lead to multiple mRNAs or multiple protein versions, which all result in proteins targeted to only one compartment. The translation of only one protein, which actually has both targeting signals in a row, is also a possibility (Fig. 2e). The decision on the actual target is then made on a different level (von Braun *et al.*, 2007). The regulatory mechanisms are manifold, be it by changing the accessibility of the specific signal or by modifying the signal by side chain modifications or by binding helper proteins. All those mechanisms could change the affinity of the signal for the import apparatus of one compartment in favor of one of the encoded targets (e.g. Waegemann and Soll (1996)). An additional possibility is an actual import into one compartment, encoded by the most N-terminal part of the signal (de Castro Silva Filho *et al.*, 1996), and then after processing, the redirection to the final destination after release of the protein from the first compartment. This has been observed for plastid proteins that are translocated via the secretory pathway, most likely by vesicle budding at the endoplasmatic reticulum (ER) and membrane fusion on the plastid side (Villarejo *et al.*, 2005) (Fig. 1).

Ambiguous targeting signals

Proteins with ambiguous targeting signals always originate from one gene and one resulting mRNA (only the targeting sequence is considered here, alternative splicing in the rest of the mRNA may occur anyway) (Fig. 2a). There is also only one protein, which has a targeting signal that can be read by the import machineries of both targets, while the underlying regulation mechanisms are very similar to the ones for twin targeted proteins: side-chain modifications of the signal peptide, protein binding to facilitate or repress import and overall affinity to the import apparatus of the compartments (e.g. Berglund *et al.* (2009) and summarized in Yogev and Pines (2011)) are used to regulate import. Those targeting peptides usually are difficult to tell apart from untypical single targeting signals, as they are set up with characteristics for both target organelles. This makes it very hard for prediction tools, to recognize them and to predict their actual localization correctly (Peeters & Small, 2001).

The occurrence of dual targeting and the detection of those mechanisms actually cast some doubt over the assumed number of proteins that are predicted or estimated for the organelles. The common assumption is that we have seen until now not the whole picture of what happens in the organelles or the other compartments that have also been shown to contain dual targeted proteins (summarized Karniely and Pines (2005)). This thought is most likely true, as it has been demonstrated that some proteins show their ability to be imported into a different compartment only under certain circumstances (Seguí-Simarro *et al.*, 2003), in certain developmental stages (Zhang *et al.*, 2010) or in certain tissues (Faraco *et al.*, 2011). The difficulties to spot those proteins at the right time in the right tissue are even more severe, as many of those proteins have one dominant localization and one, where they are present only at a very low level, quite often at or even below the detection limit of many methods (Regev-Rudzki & Pines, 2007). This eclipsed distribution does by far not mean that the localization with lower abundance represents the less important function. This second function might for example be a purely regulative one, where a low amount of protein can start a cascade of signals, e.g. the synchronization of organelle proliferation in the case of plastid and mitochondrial proteins or a pathogen response signal. *In silico* prediction tools might give hints on a second localization before experiments are started. With the knowledge of a potential second localization within the cell, more sensitive experiments might be conducted or the results of several approaches combined (Millar *et al.*, 2009). As mentioned before, not only dual targeting to the organelles is a common phenomenon, but also dual targeting to other compartments as e.g. one organelle and the nucleus. In a study on rice and *A. thaliana* it was shown that a significant number of eukaryotic transcription factors in monocotyledonous and dicotyledonous plants have a putative N-terminal targeting signal to one of the two organelles (Schwacke *et al.*, 2007). Those transcription factors are considered to be predominantly within the nucleus, so the *in vivo* confirmation of the predicted localization for more of those transcription factors in the organelles and the nucleus would make those proteins perfect mediators of anterograde and/or retrograde signaling.

ANTEROGRADE AND RETROGRADE SIGNALING

The two terms anterograde and retrograde signaling describe the direction of information conduction between nucleus and organelles. Anterograde is used for signaling from the nucleus to the organelles, while retrograde describes the signals that are sent in the other direction. As the means of anterograde signaling mediated by proteins are most likely very similar as targeting mechanisms utilized to send proteins to the organelles by the nucleus, the focus here will be mainly on retrograde signaling. There are several factors in discussion that are suggested to play a role in retrograde signaling. Amongst them are components of the tetrapyrrol pathway (GUN2-5; Mochizuki *et al.* (2001); Larkin *et al.* (2003); Strand *et al.* (2003)) and a pentatricpeptide protein (GUN1 Cottage *et al.* (2008)). The GUN-mutants (genomes uncoupled) were discovered in a mutant screen for transcription of the nuclear-encoded LHCb- (light harvesting complex b) and RBCS- (**Ru**B**i**s**C**o **s**mall subunit) genes after destruction of the plastids (Susek *et al.*, 1993). In wildtype plants, the destruction of the

plastids leads to a stop in transcription of genes that encode plastid proteins. The following assumption was that the mutation must have hit a signal that tells the nucleus to stop with the transcription. In those five mutants the affected genes belonged mainly to the tertrapyrrol pathway. One of the top candidates for a retrograde signal was magnesium protoporphyrin IX (Mg proto IX). To what extent the tetrapyrrol pathway or, for a fact, any other candidate for retrograde signaling actually transmits a signal to the nucleus is still unclear. None of the protein candidates or any other candidate molecule has been shown to actually leave the plastid, enter the nucleus and change gene expression (reviewed by Pfannschmidt (2010)). In total, several other modulators and proposed models exist, amongst them an involvement of different plant hormones, mainly abscisic acid (Shen *et al.*, 2006), Ca^{2+} (Weinl *et al.*, 2008) or metabolic pathways like carotenoid synthesis pathways (Chamovitz *et al.*, 1991) as well as combinations or networks of all proposed signals. Another model suggests that the mitochondria are the actual sensors and transmitters to the nucleus, as they sense the shift from autotrophy to heterotrophy when plastids die. The nucleus then "concludes" from that shift that the plastids are not functioning as expected and reduces transcription of plastid genes (Pfannschmidt, 2010). All those theories are in discussion right now and for all of them one can find evidence. But maybe they do not account for one fact strongly enough: retrograde signaling does not only involve "life and death" signaling from the organelles but also involves more subtle changes in the status of the organelles. In 2009, a transcription factor based model for dual targeting was suggested by Krause and Krupinska (2009). This was based on the analysis of transcription factor targeting signal predictions in rice and *A. thaliana* (Schwacke *et al.*, 2007) and on the closer analysis of AtWHY1 (Whirly 1), which is a transcription factor with dual localization in the nucleus and the plastid (Krause *et al.*, 2005). Transcription factors with dual localization could actually directly influence gene expression in the nucleus after a release from the organelle and so transmit the signal very directly.

GENE EXPRESSION IN CHLOROPLASTS OF ANGIOSPERMS

While the plastid genome still shares characteristics with prokaryotic genome organization, transcription and translation have clear mixed origin (summarized in Stern *et al.* (1997)). The plastid genome is organized in so called nucleoids, which are attached to the membranes of the plastid. Depending on the species and on the developmental status of the plastid they can be found as one single nucleoid or a dispersed pattern of smaller nucleoids (summarized inKuroiwa (1991)). Nucleoids are always attached to membranes, but not always to the same. In some species they move upon development from the envelope to the thylakoids (Sato *et al.*, 1997), in others they stay at the envelope membrane (Selldén & Leech, 1981). Several proteins have been described to anchor the DNA to the membranes. One of them is the PEND (**p**lastid **e**nvelope **n**ucleotide **b**inding) protein (Sato *et al.*, 1993; Terasawa & Sato, 2005). But nucleoids do not only contain membrane anchoring proteins or other structural proteins and DNA, but also the components of the transcriptional apparatus. It is possible to isolate the nucleoids and maintaining and enriching the transcriptional activity. This isolated

association of DNA and transcriptional activity has been named the plastid transcriptionally active chromosome (TAC). The transcriptional activity in plastids can be further divided into a soluble RNA polymerase (sRNAP – soluble RNA-polymerase) associated fraction and the DNA-bound TAC. This TAC again can be further purified and enriched to reduce method-induced bias of unspecific proteins (Krause & Krupinska, 2000). After transcription, plastid mRNAs are processed through splicing, 3' maturation and C -> U editing (summarized in Stern *et al.* (2010)), which are processes that also are partly associated with the nucleoids. The mature mRNAs are extremely stable, and plastid translation is amongst other mechanisms regulated by the presence of those mRNA pools, which accumulate within ribonucleoprotein complexes (Hayes *et al.*, 1996). Interestingly, all steps in mRNA storage, maturation, stabilization and translation seem to be organized in high molecular weight structures, similar to the nucleoids and the ribonucleoprotein complexes. They were separately isolated and analyzed in a proteomics approaches on megadalton-complexes from *A. thaliana* plastids (Nakamura *et al.*, 2004; Olinares *et al.*, 2010). This indicates a membrane independent compartmentalization within the plastid, not only for metabolic processes but also for processes related to gene expression and protein abundance within the stroma. This is comparable to the compartmentalization observed within the cell nucleus and its nucleolus and the export of mRNAs into the cytoplasm for translational processes within similar nucleoprotein complexes. The majority of the proteins needed for plastid RNA-maturation and translation are encoded by the nucleus and have to be imported into the plastid and are therefore optimal direct regulators of plastid functions.

DNA-/RNA-BINDING PROTEINS WITH DUAL LOCALIZATION IN THE NUCLEUS AND THE ORGANELLES

The majority of RNA-binding proteins described to be localized to the plastid are encoded by the nucleus. Those proteins belong to different families, amongst the biggest ones the PPR (pentatricopeptid repeat containing) proteins and proteins with an RRM (RNA-recognition motif). We have focused on three different kinds of protein (families) which are involved in RNA-related processes in the plastids and therefore are putative mediators for anterograde signaling. However, these proteins are also interesting for potential retrograde signaling, as they all have a predicted or described nuclear localization.

The Whirly transcription factor family in organelles and the nucleus

We published a review on the structures and functions of the Whirly transcription factor family in the year 2009 ((Krause *et al.*, 2009), paper II). Hence, in this paragraph only the most relevant points for potential anterograde and retrograde functions of the Whirly proteins are summarized together with what was described after the publication of that review. It has been shown in potato and later confirmed for *A. thaliana* that WHY1 is active in plant defense response and activates nuclear gene expression (Desveaux *et al.*, 2002; Desveaux *et al.*, 2004). In addition, it has been shown that AtWHY1 binds to telomeres (Yoo *et al.*, 2007b) and that HvWHY1 is present in the nucleus and the plastids of the same cell in barley (*Hordeum vulgare*), showing tetramerization in the nucleus, but not in the plastids

(Grabowski *et al.*, 2008). In *A. thaliana*, three Whirly proteins have been described: AtWHY1, ATWHy2 and AtWHY3. While AtWHY2 is a mitochondrial protein, the other two are described as plastid localized (Krause *et al.*, 2005). At WHY1 and AtWHY3 are very similar and seem to either cooperate or compensate for each other. This is indicated by the fact that the single knock-out plants (KO) for AtWHY1 or AtWHY3 do not show a strong phenotype, but the double KO shows a variegated phenotype, which is inherited by the maternal side, suggesting a plastid effect. This effect was pinpointed to a decrease in cpDNA (chloroplast DNA) stability, characterized by a higher rate of recombination. The same results were obtained by analyzing the maize KO of ZmWHY1 (maize contains only one plastid localized Whirly protein). The mechanism behind might be the binding of the Whirly proteins to ssDNA (single stranded DNA) and by that, protecting it from illegitimate recombination (Marechal *et al.*, 2009). This model was supported by a study on double-strand-break-induced micro-homology-mediated DNA-rearrangements in the double KO background (Cappadocia *et al.*, 2010). In the same study, they came to the conclusion, that Whirly proteins rather bind ssDNA than RNA, even though the affinity to RNA is given. Another study in barley came to more or less the opposite conclusion. The association of HvWHY1 to the TAC (transcriptionally active chromosome) was demonstrated to be rather RNA-mediated than DNA-mediated, and a binding of HvWHY1 to intron containing mRNAs was demonstrated (Melonek *et al.*, 2010).

In 2009, a novel function of AtWHY1 and 3 as repressors was published. AtWHY1 and AtWHY3 interact within a complex named KBF1-complex (KPRE- (At**K**inesin-**p**rotein **r**elated **e**lement)-binding factor 1) in *A. thaliana*. This factor represses the transcription of the AtKP1-gene (Xiong et al 2009). (Xiong *et al.*, 2009). Furthermore, AtWHY3 was detected in a high resolution microarray analysis to be repressed in leaf senescence as the only member of the family in *A. thaliana* in the whole study (Breeze *et al.*, 2011). This is interesting, as in barley, a function of HvWHY1 was described as influencing the senescence (Melonek *et al.*, 2010). All in all, a function for all plastidic Whirly proteins has been shown not only in the organelle but also in the nucleus. Both detected forms were actually of the same size, corresponding to the mature protein without targeting signal (Grabowski *et al.*, 2008). This year, a study was conducted to find out if the protein is exported from the plastid back to the nucleus or if a different mechanism causes this presence of the mature version of AtWHY1 in the nucleus. In transplastomic tobacco plants containing the mature form of AtWHY1 with an HA-tag, AtWHY1 was detected in the nucleus. This suggests that AtWHY1 is exported by a not yet described mechanism from the plastids induced by a not yet described signal. To exclude the possibility that this was due to an artifact, transcription of pathogen response related target genes of AtWHY1 (PR1 and PR2) was analyzed and showed upregulation compared to wildtype and so supported the specificity of the signal (Isemer *et al.*, 2012).

A highly conserved prokaryotic translation elongation factor as actor in diverse plant functions: AtEF-Tu

Not only have the two plastid localized Whirly proteins been detected in nucleoid fractions in several studies, but also AtEF-Tu (**e**longation **f**actor **t**hermo-**u**nstable) (Phinney and Thelen (2005); Majeran *et al.* (2012) and others). EF-Tu is an originally prokaryotic translation elongation factor whose gene (*tufA*) was transferred from the organelles to the nucleus in higher plants, while it remained in the plastid genome in algae (Watson & Surzycki, 1982; Baldauf & Palmer, 1990). The plant protein is still so conserved that it can transfer heat tolerance to *E. coli* cells (Moriarty *et al.*, 2002). But as the first part of the name already suggests, one of its roles in *E. coli* is the elongation of translation by binding to aminoacyl-tRNAs and GTP, mediating translation elongation by providing P_i (inorganic phosphate) through the hydrolysis of GTP to GDP and then being released from the tRNA to facilitate the creation of a peptide bond between the adjacent aminoacyl-tRNAs (Blanchard *et al.*, 2004). It fulfills that function in bacteria and also in plant organelles, but this is only the predominant function. EF-Tu is also able to activate transcription in *E. coli* (Vijgenboom *et al.*, 1988). In plants, it is involved in plastid protein synthesis (Tiboni *et al.*, 1978), stress transduction (Singh *et al.*, 2004) and transferring heat tolerance via an assumed chaperone-like activity to proteins (Bhadula *et al.*, 2001; Rao *et al.*, 2004; Ristic *et al.*, 2008). Furthermore, it has been detected in several proteomics studies within the nucleus, where they were detected in response to cold shock (Bae *et al.*, 2003) or in the nucleolus (Pendle *et al.*, 2005). This dual localization makes AtEF-Tu an interesting candidate for a possible mediator of nucleus/plastid communication, especially as it was detected in temperature response reactions in both compartments.

The AtcpRNP family: antagonists or collaborators of AtWHY1?

The protein family actually got our attention because of the described role in the nucleus: in potato, a protein named SEBF (**s**ilencing **e**lement **b**inding **f**actor) was isolated that acted as a repressor on the exact same target that StWHY1 activated. The binding of SEBF was in the same promoter region, some base pairs downstream of the binding site for StWHY1, the ERE (elicitor response element). This protein was identified as homolog to RNA-binding proteins, acted as ssDNA binding protein and TargetP (Emanuelsson *et al.*, 2000) predicted a signal for plastid localization. Nevertheless, it has been shown to act in the nucleus of potato as an antagonist to StWHY1 (Boyle & Brisson, 2001). In *A. thaliana*, the closest homolog to SEBF is AtCP29B, which is a member of the AtcpRNP-family (Ruwe *et al.*, 2011). In *Arabidopsis thaliana*, the cpRNP family has at least eight members, while in a recent publication by Ruwe *et al.* (2011) two new members were introduced. These plastid proteins are characterized by two conserved RNA-recognition motifs (RRM-domain; PFAM00076), which are separated by a glycine stretch. They also contain an N-terminal acidic domain which is believed to play a role in protein-protein interactions (Bar-Zvi *et al.*, 1992). As none of those proteins seems to have a catalytic domain but nevertheless affect mRNA-stability and maturation by binding to intron-containing tRNAs and mRNA (Nakamura et al. 1999) and assisting in C → U editing (Hirose & Sugiura, 2001), they are believed to play a role as

mediating and regulating factors for other RNA binding proteins like the PPR (pentatricopeptide repeat) proteins (Tillich *et al.*, 2010). But the similarity of the family to WHY1 does not stop with the binding to intron-containing mRNAs or the acting as transcriptional regulator in potato for plant defense. One of the best characterized members of the family, AtCP31A, has been shown under the name STEP1 to bind telomeres (Kwon & Chung, 2004; Yoo *et al.*, 2010), like it was shown for AtWHY1, too (Yoo *et al.*, 2007a). STEP1 seems to be either a splice variant or the mature form of AtCP31A. At that time, the supported explanation was the splice variant, as in a study several years earlier nine different mRNAs (Cheng *et al.*, 1994) were described. However, the annotation in TAIR (www.arabidopsis.org) shows only one splice variant, even though nine cDNAs are also annotated in the genome browser. The same version of the protein was described earlier to bind to the figwort mosaic virus (Didier & Klee, 1992). They found the protein to be expressed in all organs, from roots to rosettes and floral tissue, while other publications report a tissue-specific expression pattern for most members of the family, with a strong overrepresentation in the green tissues (Cheng *et al.* (1994); Ohta *et al.* (1995) and others). Light and developmental regulation of expression has also been proposed for some members of the family by several groups (Li & Sugiura, 1990; Mieszczak *et al.*, 1992; Churin *et al.*, 1999) and others). In *H. vulgare*, they were able to show that there is not only a general developmental regulation for expression of three members of the HvcpRNP family but also that a light independent plastid factor is needed for transcription activation for one of the genes (CP31AHv) (Churin *et al.*, 1999).

The involvement of the cpRNPs has for a long time been assumed to be rather unspecific as no specific DNA or RNA binding motif could be identified (Nakamura *et al.*, 2004). But in a study on the two AtCP31-proteins (CP31A and B), a recognition motif for plastid RNA binding was suggested for its function in C→U editing. Furthermore, a combinatorial mode of action for those two proteins was suggested as they did show a combined phenotype when both proteins were knocked out (Tillich *et al.*, 2010). If this combinatorial way of AtcpRNP-activity proves to be true and the phenotype was not observed due to a complementary mode of action of the two proteins, this is a very flexible method to fine tune the protein abundance on a post-transcriptional level. If all combinations of family members are taken into account, there are already 100 specific regulation possibilities even with the most cautious estimate, assuming that only two proteins are involved in one regulation step and no other RNA-binding proteins are involved. This would easily more than cover all 28 editing sites in the *A. thaliana* plastid transcriptome (Tillich *et al.*, 2005). Taken together, the AtcpRNPs are not only possible candidates for anterograde signaling but might also be involved in retrograde signaling due to their dual targeting properties similar to AtWHY1.

The studies on the cpRNPs and also on the Whirly proteins were conducted in different organisms, with sometimes contradictory results (Cappadocia *et al.*, 2010; Melonek *et al.*, 2010). So the raising question is, whether the proteins actually act the same way in all seed plants or whether there actually are functional and/or localization differences between the

different organisms, as at least the targeting signals have an extremely low degree of conservation even for AtEF-Tu, which is the most conserved protein in our analysis.

II. Aims of the Study

To elucidate the distribution of proteins within the cell and the communication between the organelles and the nucleus and so contributing to answer the two big questions (see pages 1-2) we aimed at clarifying the following points:

- How big is the influence of ambiguous targeting to plastids and the mitochondria?
- Are nuclear-encoded plastid RNA binding proteins potential mediators of communication between nucleus and organelles?
- Are plastid RNA binding proteins involved in more than one function?
- What is a potential regulatory mechanism for AtWHY1 in its role in anterograde or retrograde signaling?
- How transferable are *in vivo* and *in silico* approaches elucidating subcellular localization between different species?

III. Summary of publications

PAPER I

Prediction of dual targeted proteins to plant organelles

We developed a prediction tool ATP (**a**mbiguous **t**argeting **p**redictor) for ambiguous dualtargeted signals to plastids and mitochondria, which employs a machine learning approach based on support vector machines to distinguish between single targeted proteins and dual-targeted proteins. The tool uses secondary and primary structure features of the first 70 amino acids (average length of N-terminal targeting signals) of a protein, to classify the protein into single targeted or dual targeted. *In silico* validation showed that ATP is able to predict dual targeting. According to the *in silico* data, we defined a cut offof 0.7 for reliable prediction for targeting to mitochondria and plastids. Whole genome predictions with ATP revealed that the in average ≥ 400 proteins (cutoff 0.7) in seed plants might be targeted to mitochondria and plastids and therefore the composition of the plastid and mitochondrial proteomes might have to be reconsidered. *In vivo* validation on a set of *Physcomitrella patens* proteins in *P. patens* confirmed the results of the *in silico* validation, also for the defined cutoff of 0.7. Below this cut-off, the results are not that reliable anymore, but still a number of dual targeted proteins are found.

This publication is partially based on work that was included in my Diploma thesis.

PAPER II

Whirly proteins as communicators between plant organelles and the nucleus?

In this article, we reviewed the current knowledge on the Whirly transcription factor family and compared the characteristics of the protein family from green algae to seed plants. The Whirly transcription factor family is characterized by a very conserved structure consisting of eight antiparallel β-sheets in groups of four, separated by an α-helix. After the last β-sheet, two further α-helices, who form a helix-loop-helix motif, are attached. Those helices are described to play a role in protein interaction. This last domain is necessary to form the name giving, whirligig structure that has been described in crystallization experiments for potato StWHY1 homo tetramers. In angiosperms, DNA binding is mediated by a highly conserved binding domain with KGKAAL-motif. All described Whirly proteins have a prediction to be targeted to mitochondria or plastids. In *A. thaliana*, two plastid localized and one mitochondria localized homolog have been described. However, one of the first publications on Whirly transcription factors (under the name p24/PBF-2) describe nuclear functions in potato with StWHY1 being an activating factor of the plant pathogen response and binding single stranded DNA (ssDNA). In further studies, a nuclear function of AtWHY1 in pathogen response and telomere binding has been described in *A. thaliana*. In barley (*Hordeum vulgare*) it was demonstrated that HvWHY1 is an inhibitor of senescence. The first role of ZmWHY1 in plastids was described in *Zea mays* (maize), where an ivory phenotype could be observed due to lack of ribosome accumulation in the plastids. It was shown that maize ZmWHY1 binds not only to plastid DNA (*in vitro* and by isolation of TAC complex) but also to intron containing plastid mRNAs. The other plastid Whirly protein in *A. thaliana*, AtWHY3 is extremely similar to AtWHY1, but no seperate functional analysis on it has been conducted.

For the mitochondrial homolog, a role in mitochondrial genome stabilization, transcription repression in mitochondria and senescence acceleration has been shown. A retrograde signaling mechanism by relocation of the AtWHY1 protein from the plastid to the nucleus is proposed.

MANUSCRIPT I

Moonlighting in plastids - translation elongation factor EF-Tu is a component of chloroplast transcriptionally active chromosomes

We isolated the TAC complex from plastids of spinach and were able to detect the translation elongation factor EF-Tu in that isolate and confirmed the association of the protein with the TAC. Sequence analysis revealed that EF-Tu contains a putative monopartite nuclear localization signal (NLS). This NLS is located directly behind the plastid targeting signal and the targeting signal actually might cover the NLS and prevent EF-Tu to enter the nucleus in its premature form. To analyze the impact of the targeting peptide, several fusion constructs were generated, that neutralized the effect of the targeting peptide. These constructs showed a distribution throughout the cytoplasm and the nucleus. The size of the AtEF-Tu-GFP fusion rules the possibility of passive import as shown for GFP alone almost certainly out. The protein with a C-terminal fusion of GFP was directed clearly to the plastids and showed a dual pattern with the major fraction diffusely distributed in the stroma, where it co-localized with AtRPS17 (ribosomal protein of the small subunit) and the stromal part of AtCP29B (plastid ribonucleoprotein). A minor fraction showed a localization in speckles, which localized close to the plastid envelope binding protein (PEND), which is a marker for nucleoids. But AtEF-Tu showed a higher degree of co-localization with the RNA-binding proteins AtWHY1 and AtCP29B. Albeit those proteins showed co-localization, no direct protein interactions could be shown in yeast-two-hybrid interaction assays.

MANUSCRIPT II

Co-Localization and interaction of ribonucleoproteins in chloroplasts of Arabidopsis thaliana.

We were able to show that four members of the AtcpRNP family (AtCP28A, AtCP29B, AtCP31A and B) are localized in two sub-plastidic compartments: diffuse in the stroma and in speckles at the membranes. Interestingly, the predominant form of localization was different for each of the analyzed proteins, even though all of them showed both localizations. We did co-localization analyses and were able to find overlapping localizations with markers for different functions and compartments within the plastid: we did not see any co-localization with PEND, which we used as nucleoid marker, while a partial overlap with AtWHY1 was given for the speckles. Overlap with the speckles of AtEF-Tu was also partial, while the diffuse stromal fraction of AtEF-Tu (manuscript I) did overlap perfectly with the tested AtcpRNPs in the same pattern as they did with AtRPS17, the marker for ribosomes. We also analyzed the interactions with yeast-two-hybrid direct interaction assays between the proteins and were able to show that AtCP28A and AtCP31B actually did interact, while no other strong interactions could be found. Nevertheless, weak interactions were found for CP29B with CP31B and CP28A.

Localization experiments with the mature version of all proteins showed nucleo-cytoplasmic localization, which suggests that a retrograde signaling by export of the mature form like for AtWHY1 could be possible.

An analysis over the composition of related proteins showed very clearly, that there are proteins with similar structure over the whole green lineage, even within red algae, but we were not able to predict plastid localization for them. Furthermore, members of the protein family with only one RRM (RNA-recognition motif; PFAM 00076) domain were found in all analyzed genomes to be targeted to all other DNA-containing compartments. A 15 amino acid stretch was analyzed further as it is considered as the binding domain and is highly conserved. We were able to show that, depending on the predicted compartment of the protein, differences in this binding domain occurred.

MANUSCRIPT III

Can Arabidopsis thaliana *read messages from* Physcomitrella patens *and* vice versa*?* – *An analysis of the conservation of targeting signals*

We modified the codebase of ATP (paper I) to adapt it to prediction of single targeting to four different classes namely plastids, mitochondria, secretory pathway and no N-terminal targeting signal. We generated a *P. patens*-specific data set for training and testing and used as a reference the TargetP data set. By that approach, we had two predictors, GTP_Pp (*P. patens*-specific **g**reen **t**argeting **p**redictor) and GTP_Ref (TargetP-trained), available for comparison of the performance and the analysis of features used to distinguish between the classes. We were able to show that, despite a small training data set, GTP_Pp performed comparable or better on a *P. patens* test set compared to a set of published prediction tools. Furthermore, we were able to show that GTP_Pp is superior in predicting a putative second localization compared to GTP_Ref and that there are significant differences between those two tools in feature usage for mitochondrial sequences. This was confirmed in several analyses on different test sets, as well as the differences in dual targeting prediction. An *in vivo* validation approach confirmed the performance of GTP_Pp as comparable to GTP_Ref. The *in vivo* localizations were conducted with proteins from *P. patens* and *A. thaliana* in *P. patens*, *A. thaliana* and *N. tabacum*. They showed one more thing clearly: heterologous systems for localization studies can be used but might not show the full picture, especially for dual targeting. This was true even for the rather closely related species *A. thaliana* and *N. tabacum*, and should be considered when experiments are planned.

MANUSCRIPT IV "WORK IN PROGRESS"

The DNA binding domain of a Whirly protein from Arabidopsis thaliana is engaged in protein translocation across the plastid envelope membrane.

We created three different versions of deletion constructs for AtWHY1: we deleted once the DNA-binding domain (AtWHY1ΔDBD), once the protein interaction domain (AtWHY1ΔPAD) and created one construct that lacked both domains (AtWHY1ΔDBD-PAD). We did co-localization experiments of all three constructs with AtWHY1 and were able to see differences in the localization pattern for all constructs compared to the unaltered version. Interestingly, AtWHY1ΔDBD seemed to be stuck in the membrane instead of being attached to the thylakoids. The deletion of the protein interaction domain had a diverse localization pattern: it either showed localization in speckles at the thylakoids or it was distributed diffusely in the stroma. The double deletion resulted in a similar picture as the deletion of the KGKAAL domain alone. Co-localizations with the AtWHY1 showed that parts of AtWHY1ΔDBD were imported while a second fraction was still stuck in the membranes, when expressed together with the complete version of the protein. The co-localizations with AtWHY1ΔPAD showed overlap in the speckles, while AtWHY1ΔDBD-PAD was also stuck in the membranes and never showed overlap with AtWHY1. Yeast-two-hybrid direct interaction studies showed a change in the ability to interact when AtWHY1 was tested against the deletions. Not surprising, the interaction was completely abolished in the deletions of the protein interaction domain, but the interaction was also disturbed when the DNA binding domain was deleted. This might be due either to a conformational change or the DNA binding domain is actually involved in protein-protein interaction or, as a third option, DNA binding mediates protein-protein interaction.

IV. Discussion and Outlook

I present in this work data that adds a puzzle piece to the big picture of plastid targeting and signaling. This might, in the long run, help to answer the two guiding questions this work was based on. In this section, the contribution of the data presented to each question is summarized and future prospects are discussed.

> *How many and which proteins are present in which compartment at what time and why?*

The research presented in paper I and manuscript III was conducted to understand and to elucidate the protein distribution within the cell further and to take a step to answering parts of this question. We were able to provide the first predictor for ambiguous targeting signals ATP (paper I). To be able to not only predict single targeting but also be able to predict N-terminally encoded, potential dual targeting signals will bring the annotation of protein distribution within the cell forward. Nevertheless, the prediction tool was based on a relatively small data set and will probably improve significantly with a growing number of experimentally confirmed dual targeted proteins and an accompanying increase in test and training data set. Nevertheless, we predicted a not expected high number of dual targeted proteins with ≥ 400 in average per predicted genome. The second prediction tool, GTP_Pp (manuscript III), was developed not only to compare the divergences between targeting signals throughout the green lineage, but also to see if a species-specific predictor actually has advantages compared to predictor trained on a mixed data set. We came to the conclusion that especially mitochondrial and dual targeting signals diverge to such a degree that a species-specific prediction tool is desirable and a species-specific experimental setup almost indispensable. However, the size of the data set was also for that prediction tool a challenge, but with the gain in experimental data, species-specific tools should be considered as the desirable goal for the future. This kind of *in silico* analysis can help to resolve questions concerning spatial distribution of proteins, but they are unfortunately not able to answer the question of temporal distribution. The temporal distribution and the regulation behind will have to be approached by experimental procedures in the lab. But our data also showed that to actually be able to detect as many dual targeted proteins as possible, homologous expression systems should be used whenever possible.

As it has been shown in the past, finding candidates that actually are involved in the intracellular communication is difficult as until now, only indirect evidence for an actual signal or regulation mechanism exists. Based on the hypothesis introduced by Krause and Krupinska (2009) we chose candidates to focus on, which fulfill the proposed criteria: we chose proteins with a function in transcriptional regulation and a putative dual localization in plastids and nucleus. Our candidates were chosen from two different protein families (AtWhirly transcription factor family: paper II and manuscript IV "work in progress"; AtcpRNPs: manuscript II) and one single protein (manuscript I: AtEF-Tu). These proteins share their localization pattern in plastids (organelles) and the nucleus, a function in RNA-related processes in plastids and a possible function in transcription regulation in the nucleus.

For one member of the Whirly transcription factor family, AtWHY1, it has already been shown that it is exported from the plastids to relocate to the nucleus and our aim was to identify the underlying mechanism. We were able to show with the AtWHY1ΔDBD construct that the DNA-binding domain of AtWHY1 affects the localization and the ability to interact with AtWHY1, while the deletion of the protein binding domain (AtWHY1ΔPAD) does at least not affect the localization in the plastid as severely. Therefore, we think that the possibility to bind DNA/RNA is crucial for regulation, either for anterograde or retrograde signaling. At the moment, we have several hypotheses on the mechanism behind the observed patterns of the deletion constructs of AtWHY1. Future work will be aimed at a refinement and validation of one of those models to maybe integrate them into a new, different and more complex model. This refinement and confirmation has to be approached under several angles: On the one hand, a comparison of the expression of nuclear genes between AtWHY1 and AtWHY1ΔDBD overexpressing plants will give hints on further direct targets of AtWHY1 in the nucleus. On the other hand, an expression profile with plants overexpressing AtWHY1 lacking its ability to build homo tetramers in comparison to the complete AtWHY1 might help to distinguish between direct and indirect effects on the nuclear gene expression, as the AtWHY1 protein has been shown to be active in homo tetramers in the nucleus while in the plastid monomers seem to be the predominant version (Grabowski *et al.*, 2008).

A third aspect will be to determine which mechanism of translocation is actually disturbed in AtWHY1ΔDBD constructs: import into the plastid or export from the plastid? So it would be possible to get closer to a decision, which one of the three hypotheses presented in figure 3 could be true:

We are planning to evaluate the importance of DNA/RNA binding for import and so for anterograde signaling by *in vitro* import assays in isolated organelles with AtWHY1ΔDBD and AtWHY1ΔPAD in comparison to AtWHY1. This will show if there is a change in import behavior depending on the possibility to bind DNA/RNA or build tetramers without the influence of protein overexpression. The influence of DNA/RNA binding is at the moment considered as the more likely option, as AtWHY1ΔPAD does not show problems with import

("work in progress"). That mRNA has to be imported into plastids has been shown before (Nicolai *et al.*, 2007) and maybe AtWHY1 is involved in that process (Fig. 3b).

Experiments similar to the ones conducted on transplastomic tobacco by Isemer *et al.* (2012) will give hints about possible changes in export dynamics for the AtWHY1ΔDBD construct compared to AtWHY1. In case we can actually confirm a change of these dynamics of AtWHY1ΔDBD from the plastids, this could lead to a preliminary model for a retrograde signaling regulation mechanism based on DNA/RNA binding. The message that could be transmitted by this mechanism would either be based on a DNA binding model (Fig. 3a) or on an RNA binding model (Fig. 3c).

The DNA binding model could be based on the findings by Marechal *et al.* (2009), who showed that AtWHY1 is necessary to facilitate the repair of double-strand-breaks and so ensuring the stability of the organellar genome. Combined with the observed localization for the AtWHY1ΔDBD construct, we suggest the following model: When double-strand-breaks happened a lot of ssDNA is present and therefore the majority of AtWHY1 is bound. Therefore it cannot be released from the plastid, while the plastid genome is unaffected, a lot of AtWHY1 is unbound and can therefore be released from the plastid. This would translate into a signal of well-being by the plastid in a constant flux of AtWHY1 to the nucleus. We have to assume for this model that a certain amount of AtWHY1 has to be present within the plastid, so that an immediate response to double-strand-breaks is possible. This existence of such a possible pool of free AtWHY1 is actually supported by the fact that Grabowski *et al.* (2008) showed that a substantial amount of HvWHY1 in barley is located to the stroma. As a consequence, the export would be triggered by the amount of free AtWHY1 in the plastid and if the accumulation is high enough, export is initiated. The observed pattern for AtWHY1ΔDBD could be explained then by a saturation of the export channels or receptors due to overexpression (Small *et al.*, 1998).

The RNA binding model could be based on the fact that HvWHY1 is, amongst other proteins, involved in mRNA maturation in plastids (Melonek *et al.*, 2010). This means that the amount of free AtWHY1 would depend on the amount of available target mRNA. The further release mechanisms would be similar to the DNA-binding model, including the pool of unbound AtWHY1 within the plastid to accommodate for dynamic processes. As a consequence, a release of unbound AtWHY1 would signal to the nucleus that only low amounts of specific transcription and mRNA processing and protein production takes place. The nucleus would get direct information on the rate of the plastid mRNA processing and could react accordingly. In case other proteins involved in mRNA-maturation and protein translation (like e.g. the AtcpRNPS or AtEF-Tu) could function through the same dual localization mediated mechanism, the nucleus could get an extremely detailed picture on what kind of transcription and how much of it takes place within the plastid at a given time point.

In what way these very simple models reflect the real situation within the organelle and which other factors actually have to be included will be shown in the future. The in this study presented possible network of overlapping and entangled multiple functions of RNA binding

proteins within the plastid might have to be considered in this scenario. So the picture of retrograde and/or anterograde signaling by plastid RNA binding proteins might be extremely complex. One group of proteins that has already been described to have several functions within the plastid in different organisms are the in this study analyzed AtcpRNPs.

Our work on this protein family (manuscript II) confirmed that they have not only one function in *A. thaliana* plastids, but also that those functions probably take place in different compartments within the plastid. cpRNPs have a function in mRNA stabilization (Nakamura *et al.*, 2001) and in C -> U editing (Tillich *et al.*, 2009), which explains the two observed patterns. The mRNAs that have to be stabilized are present in the stroma (Nakamura *et al.*, 2001), while the editing is loosely associated with the TAC (Fig. 4), where some members of the AtcpRNP family actually also have been detected in a mass spectrometry analysis of nucleoids (Majeran *et al.*, 2012). We found strong protein-protein interactions between two members of the family (AtCP28A and AtCP31B), and some weak interaction (AtCP29B with AtCP28A and AtCP31B), but the majority did not interact. To gain a more complete picture on possible interactions, all plastid localized RNP proteins will have to be tested. It would also be interesting to see, in which compartment the interactions actually take place and if they are necessary for the function in the nucleus or for the function in the plastids. A possible role in retrograde signaling could not be shown. Nevertheless, a link between RNA-editing seems to exist with GUN1 being a PPR protein, a family that has been described to be crucial in RNA editing in organelles (Schmitz-Linneweber & Small, 2008). In addition, a potential interaction of retrograde signaling defects and RNA-editing has been postulated recently (Kakizaki *et al.*, 2012). An interesting approach to elucidate this potential mechanism further might be the analysis of the actual RNA binding motif, to see if congruence with the mode of action exists for AtWHY1 and the AtcpRNPs. The last two amino acids of the 15 amino acid stretch of the RNA-binding domain seem to be a lot more conserved in plastids than in proteins with the same domain targeted to other compartments (manuscript II). It would be interesting to compare the targeting behavior when those domains are mutated either completely or to a “mitochondria-type motif”. Also the deletion of those motifs or parts of it could give a hint on possible nuclear targets in a similar approach as suggested for AtWHY1ΔDBD. In case changes in nuclear gene expression could be observed for any of those deletions or mutations it would be a hint on a possibly common regulation mechanism between mitochondria and plastids via RNA-binding proteins. This might already be assumed by a look at the composition of the RNA-binding protein families. All of them seem to have plastid and mitochondrial members, which have at least similar functions in both compartments, as it has been shown for the Whirly family. AtWHY2 is a mitochondrial protein, where it has functions in genome stability maintenance (Marechal *et al.*, 2008). A group of RNA-binding proteins containing a RRM motif has also been described in mitochondria, even though their function still has to be determined (Vermel *et al.*, 2002). However, they are discussed to fulfill a similar function as their plastidic counterparts (Takenaka *et al.*, 2008).

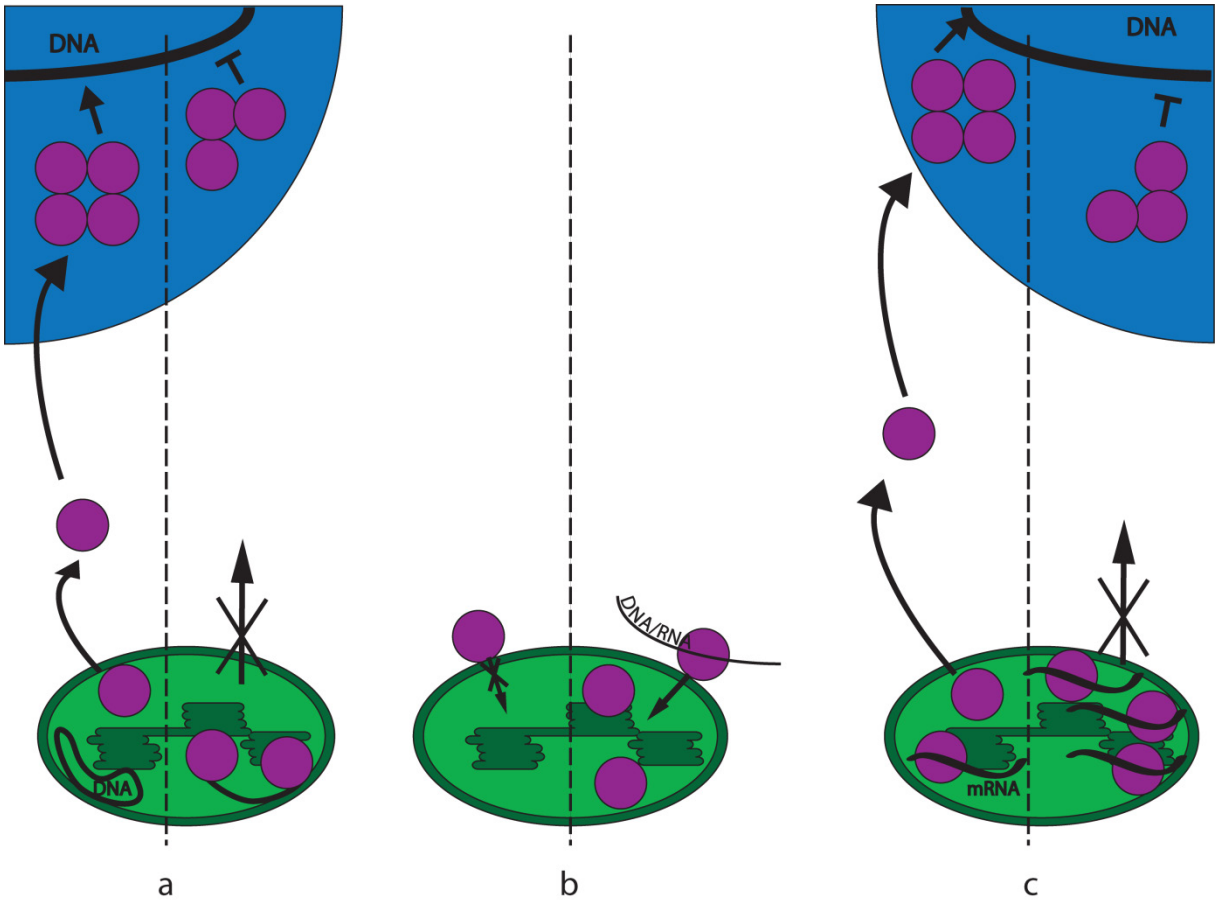


Figure 3: Three putative models for AtWHY1-mediated signaling as suggested by the deletion of the DNA-binding domain are shown.
Model (a) shows, how the retrograde regulation by a DNA-binding mechanism could work: When only very few double-strand-breaks occur in the plastid (green), a lot of free AtWHY1 (pink) is available (left). The unbound AtWHY1 is then exported and imported into the nucleus (blue). Tetramers can then start transcription. If AtWHY1 is bound, nothing is exported (right).
Model (b) shows a possible explanation for the observed pattern in AtWHY1ΔDBD due to inefficient import. AtWHY1 might only be imported, when DNA or RNA is bound (right).
Model (c) shows a second, putative retrograde mechanism due to mRNA binding. When the transcript pool within the plastid is big, only little AtWHY1 is free (right), while when transcription within the plastid is very low, the mRNA pool would be small and a lot of AtWHY1 would be free. Export from the plastid and import into the nucleus would be possible (right).

The last protein we analyzed was AtEF-Tu. It showed partial overlap with members of both other families in the plastid and might be involved in TAC related functions and not only translation elongation in the plastid (manuscript II). AtEF-Tu is the only candidate, which has a putative NLS, even though it might be concealed in the pre-mature protein (51.6 kDa). AtEF-Tu is actually also the biggest protein from our set of candidates with a size of 45 kDa for the mature protein. This is above the nuclear pore exclusion size (~50 kDa) when fused to a fluorescence protein (mature EF-Tu – GFP: ~ 71 kDa) (Macara, 2001), in contrast to the size of the mature fusion proteins from the other two families. Nevertheless, all mature

versions of the tested proteins have been detected in the nucleus when expressed in tobacco protoplasts. For plants EF-Tu proteins a nuclear function or a function as a transcriptional regulator has still to be assessed, even though it has been shown in *E. coli* that it can influence gene expression (Vijgenboom *et al.*, 1988). Yet, it has been detected in several nuclear protein extracts (amongst others Pendle *et al.* (2005)). It will be interesting to see, if an actual function of this protein will be detected in the nucleus and if it really isinvolved in retrograde signaling, too.

Nevertheless, we can propose a model of action for the AtWhirly proteins together with the AtcpRNPs and AtEF-Tu in build on the markers in use (Fig. 4) within the plastids, based on our co-localization data. We were able to show, that all proteins show partial overlap and their localizations seem to "melt" into each other. This observation actually matches the described functions within the plastid for those proteins, also the representation of roughly two areas within the plastid where they accumulate.

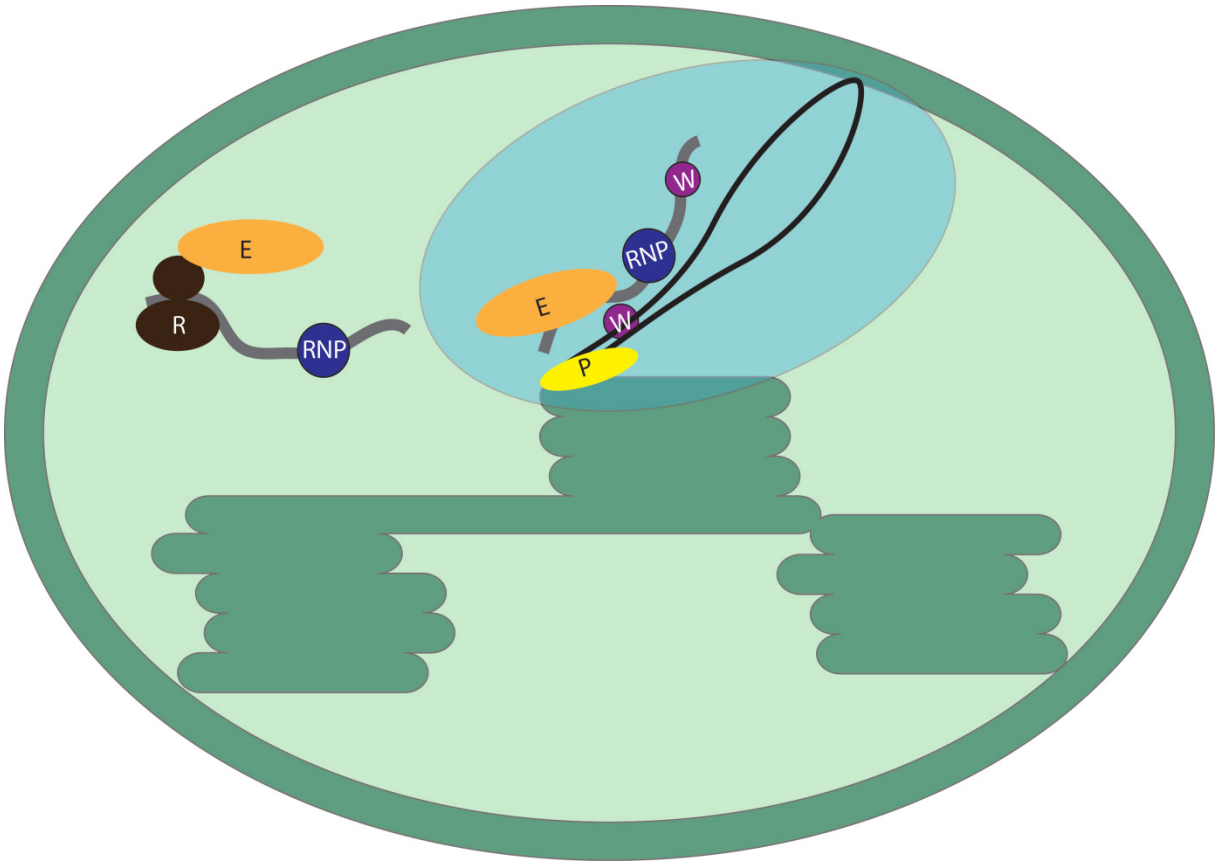


Figure 4: Depicted is a model of the described RNA and DNA binding proteins within the plastid related to their proposed functions. DNA is shown in black, RNA in grey. The stroma is light green while thylakoids and intramembrane space is dark green. One of the nucleoids is exemplarily indicated as a blue circle. The markers PEND (P) and AtRPS17 (within the ribosomes R) are shown in yellow and brown, respectively. AtEF-Tu (E) is shown in orange, on the one hand partially overlapping with PEND, AtWHY1 and AtCP29B. All AtcpRNPs are depicted combined as RNP (dark blue). On the other hand AtEF-Tu is overlapping in the stroma with the ribosomes and the stromal fraction of the AtcpRNPs, which itself do also co-localize with AtRPS17. AtWHY1 (W) is shown to bind on the one hand DNA and partially

overlaps with PEND (Melonek et al 2010), while the overlap with the AtcpRNPs and AtEF-Tu is probably due to RNA binding.

V. Conclusion

Taken together, we can say that even though we are still far away from answering the two questions completely, we made an approach to understand parts of the underlying mechanisms a bit better. We showed the potential impact of dual targeting to mitochondria and plastids on the protein distribution within the whole cell and also the importance of species-specific signals on protein targeting, which should be considered when starting experiments. Based on our data we suggest conducting experiments, whenever technically possible in the homologous system to be able to detect dual targeting. As dual targeting is an important regulatory mechanism, not only possibly in retrograde and anterograde signaling, but also in adaption to environmental and developmental changes, it is important to fully understand the localization potential of as many proteins as possible.

To further elucidate the regulatory potential behind dual targeting, we showed with the deletions of the AtWHY1 protein that a nucleic acid dependent regulation mechanism of targeting might exist and that, in combination with the work by Isemer *et al.* (2012), this might even be a retrograde signaling mechanism. If this could be confirmed in the future, this would be the first time not only the existence of a plastid factor mediating retrograde signaling is shown directely, but also the possible mechanism is uncovered. The fact that AtcpRNPs seem to be able to interact and also have more than one function in the plastid opens up for some new perspectives for a mechanism to fine tune regulation. Until now, this is mainly anterograde regulation, but future research will show, if the suggested retrograde mechanisms are true for this protein family, too. The fact that AtEF-Tu might also be a candidate in this setup was not expected. And after all, it first has to be proven to be a transcriptional regulator in the nucleus to actually qualify for a direct messenger of retrograde signals. Possibly, a different function within the nucleus will be found to support its role in retrograde signaling. Nevertheless, its role in anterograde signaling seems clear, as it is not only regulating translation elongation, but might also be involved in RNA-maturation or other RNA-related processes within the nucleoid.

We show here, that anterograde and retrograde signaling mediated by dual targeted RNA binding proteins is a mechanism that should be taken into account when analyzing nucleus-to-organelle communication.

VI. References

Arabidopsis Genome I. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**(6814): 796-815.

Bae MS, Cho EJ, Choi EY, Park OK. 2003. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant Journal* **36**(5): 652-663.

Bahaji A, Ovecka M, Barany I, Risueno MC, Munoz FJ, Baroja-Fernandez E, Montero M, Li J, Hidalgo M, Sesma MT, Ezquer I, Testillano PS, Pozueta-Romero J. 2011. Dual targeting to mitochondria and plastids of AtBT1 and ZmBT1, two members of the mitochondrial carrier family. *Plant and Cell Physiology* **52**(4): 597-609.

Baldauf SL, Palmer JD. 1990. Evolutionary transfer of the chloroplast tufA gene to the nucleus. *Nature* **344**(6263): 262-265.

Bar-Zvi D, Shagan T, Schindler U, Cashmore AR. 1992. RNP-T, a ribonucleoprotein from Arabidopsis thaliana, contains two RNP-80 motifs and a novel acidic repeat arranged in an alpha-helix conformation. *Plant Mol Biol* **20**(5): 833-838.

Berglund AK, Pujol C, Duchene AM, Glaser E. 2009. Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. *J Mol Biol* **393**(4): 803-814.

Bhadula SK, Elthon TE, Habben JE, Helentjaris TG, Jiao S, Ristic Z. 2001. Heat-stress induced synthesis of chloroplast protein synthesis elongation factor (EF-Tu) in a heat-tolerant maize line. *Planta* **212**(3): 359-366.

Bhattacharya D, Archibald JM, Weber AP, Reyes-Prieto A. 2007. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **29**(12): 1239-1246.

Blanchard SC, Gonzalez RL, Kim HD, Chu S, Puglisi JD. 2004. tRNA selection and kinetic proofreading in translation. *Nat Struct Mol Biol* **11**(10): 1008-1014.

Bowsher CG, Tobin AK. 2001. Compartmentation of metabolism within mitochondria and plastids. *Journal of experimental botany* **52**(356): 513-527.

Boyle B, Brisson N. 2001. Repression of the defense gene PR-10a by the single-stranded DNA binding protein SEBF. *Plant Cell* **13**(11): 2525-2537.

Breeze E, Harrison E, McHattie S, Hughes L, Hickman R, Hill C, Kiddle S, Kim YS, Penfold CA, Jenkins D, Zhang C, Morris K, Jenner C, Jackson S, Thomas B, Tabrett A, Legaie R, Moore JD, Wild DL, Ott S, Rand D, Beynon J, Denby K, Mead A, Buchanan-Wollaston V. 2011. High-resolution temporal profiling of transcripts during Arabidopsis leaf senescence reveals a distinct chronology of processes and regulation. *Plant Cell* **23**(3): 873-894.

Briesemeister S, Blum T, Brady S, Lam Y, Kohlbacher O, Shatkay H. 2009. SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. *Journal of proteome research* **8**(11): 5363-5366.

Bruce BD. 2001. The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim Biophys Acta* **1541**(1-2): 2-21.

Cappadocia L, Marechal A, Parent JS, Lepage E, Sygusch J, Brisson N. 2010. Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. *Plant Cell* **22**(6): 1849-1867.

Chabregas SM, Luche DD, Van Sluys MA, Menck CF, Silva-Filho MC. 2003. Differential usage of two in-frame translational start codons regulates subcellular localization of Arabidopsis thaliana THI1. *J Cell Sci* **116**(Pt 2): 285-291.

Chamovitz D, Pecker I, Hirschberg J. 1991. The molecular basis of resistance to the herbicide norflurazon. *Plant Mol Biol* **16**(6): 967-974.

Cheng SH, Cline K, DeLisle AJ. 1994. An Arabidopsis chloroplast RNA-binding protein gene encodes multiple mRNAs with different 5' ends. *Plant Physiol* **106**(1): 303-311.

Christensen AC, Lyznik A, Mohammed S, Elowsky CG, Elo A, Yule R, Mackenzie SA. 2005. Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons. *Plant Cell* **17**(10): 2805-2816.

Churin Y, Hess WR, Borner T. 1999. Cloning and characterization of three cDNAs encoding chloroplast RNA-binding proteins from barley (Hordeum vulgare L.):

differential regulation of expression by light and plastid development. *Curr Genet* **36**(3): 173-181.

Cleary SP, Tan FC, Nakrieko KA, Thompson SJ, Mullineaux PM, Creissen GP, von Stedingk E, Glaser E, Smith AG, Robinson C. 2002. Isolated plant mitochondria import chloroplast precursor proteins in vitro with the same efficiency as chloroplasts. *J Biol Chem* **277**(7): 5562-5569.

Cottage AJ, Mott EK, Wang J-H, Sullivan JA, MacLean D, Tran L, Choy M-K, Newell C, Kavanagh TA, Aspinall S, Gray JC 2008. GUN1 (GENOMES UNCOUPLED1) Encodes a Pentatricopeptide Repeat (PPR) Protein Involved in Plastid Protein Synthesis-Responsive Retrograde Signaling to the NucleusPhotosynthesis. Energy from the Sun. In: Allen JF, Gantt E, Golbeck JH, Osmond B eds.: Springer Netherlands, 1202-1205.

Creissen G, Reynolds H, Xue Y, Mullineaux P. 1995. Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. *Plant J* **8**(2): 167-175.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A Sequence Logo Generator. *Genome Research* **14**(6): 1188-1190.

Dabney-Smith C, van Den Wijngaard PW, Treece Y, Vredenberg WJ, Bruce BD. 1999. The C terminus of a chloroplast precursor modulates its interaction with the translocation apparatus and PIRAC. *J Biol Chem* **274**(45): 32351-32359.

de Castro Silva Filho M, Chaumont F, Leterme S, Boutry M. 1996. Mitochondrial and chloroplast targeting sequences in tandem modify protein import specificity in plant organelles. *Plant Mol Biol* **30**(4): 769-780.

Desveaux D, Allard J, Brisson N, Sygusch J. 2002. A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol* **9**(7): 512-517.

Desveaux D, Despres C, Joyeux A, Subramaniam R, Brisson N. 2000. PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* **12**(8): 1477-1489.

Desveaux D, Subramaniam R, Despres C, Mess JN, Levesque C, Fobert PR, Dangl JL, Brisson N. 2004. A "whirly" transcription factor is required for salicylic acid-dependent disease resistance in Arabidopsis. *Developmental cell* **6**(2): 229-240.

Didier DK, Klee HJ. 1992. Identification of an Arabidopsis DNA-binding protein with homology to nucleolin. *Plant Mol Biol* **18**(5): 977-979.

Dinkins RD, Majee SM, Nayak NR, Martin D, Xu Q, Belcastro MP, Houtz RL, Beach CM, Downie AB. 2008. Changing transcriptional initiation sites and alternative 5'- and 3'-splice site selection of the first intron deploys Arabidopsis protein isoaspartyl methyltransferase2 variants to different subcellular compartments. *Plant Journal* **55**(1): 1-13.

Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the molecular machines for protein import into mitochondria. *Science* **313**(5785): 314-318.

Duchene AM, Giritch A, Hoffmann B, Cognat V, Lancelin D, Peeters NM, Zaepfel M, Marechal-Drouard L, Small ID. 2005. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **102**(45): 16484-16489.

Dyall SD, Brown MT, Johnson PJ. 2004. Ancient invasions: from endosymbionts to organelles. *Science* **304**(5668): 253-257.

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal* **45**(4): 616-629.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**(4): 1005-1016.

Faraco M, Di Sansebastiano GP, Spelt K, Koes RE, Quattrocchio FM. 2011. One protoplast is not the other! *Plant Physiol* **156**(2): 474-478.

Franzen LG, Rochaix JD, von Heijne G. 1990. Chloroplast transit peptides from the green alga Chlamydomonas reinhardtii share features with both mitochondrial and higher plant chloroplast presequences. *FEBS letters* **260**(2): 165-168.

Fuss J, Liegmann O, Krause K, Rensing SA. 2012. Can Arabidopsis thaliana read the messages from Physcomitrella patens and vice versa? - An analysis on the conservation of targeting signals. *manuscript submitted to New Phytologist*.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* **296**(5565): 92-100.

Gonzalez-Lamothe R, Boyle P, Dulude A, Roy V, Lezin-Doumbou C, Kaur GS, Bouarab K, Despres C, Brisson N. 2008. The transcriptional activator Pti4 is required for the recruitment of a repressosome nucleated by repressor SEBF at the potato PR-10a gene. *Plant Cell* **20**(11): 3136-3147.

Grabowski E, Miao Y, Mulisch M, Krupinska K. 2008. Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell. *Plant Physiol* **147**(4): 1800-1804.

Gross J, Bhattacharya D. 2009. Mitochondrial and plastid evolution in eukaryotes: an outsiders' perspective. *Nat Rev Genet* **10**(7): 495-505.

Hayes R, Kudla J, Schuster G, Gabay L, Maliga P, Gruissem W. 1996. Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. *Embo Journal* **15**(5): 1132-1141.

Hedtke B, Borner T, Weihe A. 1997. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* **277**(5327): 809-811.

Henikoff JG, Greene EA, Pietrokovski S, Henikoff S. 2000. Increased coverage of protein families with the blocks database servers. *Nucleic Acids Res* **28**(1): 228-230.

Herrmann U, Fuss J, Krupinska K, Krause K. 2012. Moonlighting in plastids - translation elongation factor EF-Tu is a component of chloroplast transcriptionally active chromosomes. *Manuscript submitted to Current Genetics*.

Hirose T, Sugiura M. 2001. Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast in vitro RNA editing system. *Embo Journal* **20**(5): 1144-1152.

Hoglund A, Donnes P, Blum T, Adolph HW, Kohlbacher O. 2006. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **22**(10): 1158-1165.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**(Web Server issue): W585-587.

Hu J, Coombes K, Morris J, Baggerly K. 2005. The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Briefings in functional genomics & proteomics* **3**(4): 322-353.

Huang J, Hack E, Thornburg RW, Myers AM. 1990. A yeast mitochondrial leader peptide functions in vivo as a dual targeting signal for both chloroplasts and mitochondria. *Plant Cell* **2**(12): 1249-1260.

Huang S, Taylor NL, Whelan J, Millar AH. 2009. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol* **150**(3): 1272-1285.

Isemer R, Mulisch M, Schafer A, Kirchner S, Koop HU, Krupinska K. 2012. Recombinant Whirly1 translocates from transplastomic chloroplasts to the nucleus. *FEBS Lett* **586**(1): 85-88.

Jacobs J, Kuck U. 2011. Function of chloroplast RNA-binding proteins. *Cell Mol Life Sci* **68**(5): 735-748.

Jeong SY, Rose A, Meier I. 2003. MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure. *Nucleic Acids Res* **31**(17): 5175-5185.

Kakizaki T, Yazu F, Nakayama K, Ito-Inaba Y, Inaba T. 2012. Plastid signalling under multiple conditions is accompanied by a common defect in RNA editing in plastids. *J Exp Bot* **63**(1): 251-260.

Karniely S, Pines O. 2005. Single translation--dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Rep* **6**(5): 420-425.
Kennedy BK. 2002. Mammalian transcription factors in yeast: strangers in a familiar land. *Nat Rev Mol Cell Biol* **3**(1): 41-49.
Krause K, Herrmann U, Fuss J, Miao Y, Krupinska K. 2009. Whirly proteins as communicators between plant organelles and the nucleus? *Endocytobiosis Cell Res.* **19**: 51-62.
Krause K, Kilbienski I, Mulisch M, Rodiger A, Schafer A, Krupinska K. 2005. DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles. *FEBS Lett* **579**(17): 3707-3712.
Krause K, Krupinska K. 2000. Molecular and functional properties of highly purified transcriptionally active chromosomes from spinach chloroplasts. *Physiologia Plantarum* **109**(2): 188-195.
Krause K, Krupinska K. 2009. Nuclear regulators with a second home in organelles. *Trends Plant Sci* **14**(4): 194-199.
Kuroiwa T. 1991. The Replication, Differentiation, and Inheritance of Plastids with Emphasis on the Concept of Organelle Nuclei. *International Review of Cytology-a Survey of Cell Biology* **128**: 1-62.
Kwon C, Chung IK. 2004. Interaction of an Arabidopsis RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. *J Biol Chem* **279**(13): 12812-12818.
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.
Larkin RM, Alonso JM, Ecker JR, Chory J. 2003. GUN4, a regulator of chlorophyll synthesis and intracellular signaling. *Science* **299**(5608): 902-906.
Li Y, Sugiura M. 1990. Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs. *The EMBO journal* **9**(10): 3059-3125.
Lingner T, Kataya AR, Antonicelli GE, Benichou A, Nilssen K, Chen XY, Siemsen T, Morgenstern B, Meinicke P, Reumann S. 2011. Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. *Plant Cell* **23**(4): 1556-1572.
Macara IG. 2001. Transport into and out of the nucleus. *Microbiol Mol Biol Rev* **65**(4): 570-594, table of contents.
Macasev D, Newbigin E, Whelan J, Lithgow T. 2000. How do plant mitochondria avoid importing chloroplast proteins? Components of the import apparatus Tom20 and Tom22 from Arabidopsis differ from their fungal counterparts. *Plant physiology* **123**(3): 811-817.
Majeran W, Friso G, Asakura Y, Qu X, Huang M, Ponnala L, Watkins KP, Barkan A, van Wijk KJ. 2012. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. *Plant Physiol* **158**(1): 156-189.
Marechal A, Parent JS, Sabar M, Veronneau-Lafortune F, Abou-Rached C, Brisson N. 2008. Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* **8**(1): 42.
Marechal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. 2009. Whirly proteins maintain plastid genome stability in Arabidopsis. *Proc Natl Acad Sci U S A* **106**(34): 14693-14698.
Martin W, Russell MJ. 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* **358**(1429): 59-83; discussion 83-55.
Melonek J, Mulisch M, Schmitz-Linneweber C, Grabowski E, Hensel G, Krupinska K. 2010. Whirly1 in chloroplasts associates with intron containing RNAs and rarely co-localizes with nucleoids. *Planta* **232**(2): 471-481.
Mieszczak M, Klahre U, Levy JH, Goodall GJ, Filipowicz W. 1992. Multiple plant RNA binding proteins identified by PCR: expression of cDNAs encoding RNA binding

proteins targeted to chloroplasts in Nicotiana plumbaginifolia. *Mol Gen Genet* **234**(3): 390-400.

Millar AH, Carrie C, Pogson B, Whelan J. 2009. Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell* **21**(6): 1625-1631.

Millar AH, Whelan J, Small I. 2006. Recent surprises in protein targeting to mitochondria and plastids. *Curr Opin Plant Biol* **9**(6): 610-615.

Mireau H, Lancelin D, Small ID. 1996. The same Arabidopsis gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. *Plant Cell* **8**(6): 1027-1039.

Mochizuki N, Brusslan JA, Larkin R, Nagatani A, Chory J. 2001. Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc Natl Acad Sci U S A* **98**(4): 2053-2058.

Mockli N, Auerbach D. 2004. Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. *BioTechniques* **36**(5): 872-876.

Moriarty T, West R, Small G, Rao D, Ristic Z. 2002. Heterologous expression of maize chloroplast protein synthesis elongation factor (EF-Tu) enhances Escherichia coli viability under heat stress. *Plant Science* **163**(6): 1075-1082.

Nakamura T, Ohta M, Sugiura M, Sugita M. 1999. Chloroplast ribonucleoproteins are associated with both mRNAs and intron-containing precursor tRNAs. *FEBS letters* **460**(3): 437-441.

Nakamura T, Ohta M, Sugiura M, Sugita M. 2001. Chloroplast ribonucleoproteins function as a stabilizing factor of ribosome-free mRNAs in the stroma. *J Biol Chem* **276**(1): 147-152.

Nakamura T, Schuster G, Sugiura M, Sugita M. 2004. Chloroplast RNA-binding and pentatricopeptide repeat proteins. *Biochem Soc Trans* **32**(Pt 4): 571-574.

Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic Metabolism in Plastids. *Annu Rev Plant Physiol Plant Mol Biol* **51**(1): 111-140.

Nicolai M, Duprat A, Sormani R, Rodriguez C, Roncato MA, Rolland N, Robaglia C. 2007. Higher plant chloroplasts import the mRNA coding for the eucaryotic translation initiation factor 4E. *FEBS Lett* **581**(21): 3921-3926.

Obara K, Sumi K, Fukuda H. 2002. The use of multiple transcription starts causes the dual targeting of Arabidopsis putative monodehydroascorbate reductase to both mitochondria and chloroplasts. *Plant and Cell Physiology* **43**(7): 697-705.

Ohta M, Sugita M, Sugiura M. 1995. Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in Arabidopsis thaliana: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. *Plant Mol Biol* **27**(3): 529-539.

Olinares PD, Ponnala L, van Wijk KJ. 2010. Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol Cell Proteomics* **9**(7): 1594-1615.

Padmasree K, Padmavathi L, Raghavendra AS. 2002. Essentiality of mitochondrial oxidative metabolism for photosynthesis: optimization of carbon assimilation and protection against photoinhibition. *Crit Rev Biochem Mol Biol* **37**(2): 71-119.

Palmer JD. 2003. The Symbiotic Birth and Spread of Plastids: How Many Times and Whodunit? *Journal of Phycology* **39**(1): 4-12.

Papenbrock J, Grimm B. 2001. Regulatory network of tetrapyrrole biosynthesis--studies of intracellular signalling involved in metabolic and developmental control of plastids. *Planta* **213**(5): 667-681.

Peeters N, Small I. 2001. Dual targeting to mitochondria and chloroplasts. *Biochim Biophys Acta* **1541**(1-2): 54-63.

Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ. 2005. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol Biol Cell* **16**(1): 260-269.

Pfannschmidt T. 2010. Plastidial retrograde signalling--a true "plastid factor" or just metabolite signatures? *Trends Plant Sci* **15**(8): 427-435.

Pfeiffer A, Kunkel T, Hiltbrunner A, Neuhaus G, Wolf I, Speth V, Adam E, Nagy F, Schafer E. 2009. A cell-free system for light-dependent nuclear import of phytochrome. *Plant Journal* **57**(4): 680-689.

Phinney BS, Thelen JJ. 2005. Proteomic characterization of a triton-insoluble fraction from chloroplasts defines a novel group of proteins associated with macromolecular structures. *J Proteome Res* **4**(2): 497-506.
Prikryl J, Watkins KP, Friso G, van Wijk KJ, Barkan A. 2008. A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* **36**(16): 5152-5165.
Rao D, Momcilovic I, Kobayashi S, Callegari E, Ristic Z. 2004. Chaperone activity of recombinant maize chloroplast protein synthesis elongation factor, EF-Tu. *Eur J Biochem* **271**(18): 3684-3692.
Regev-Rudzki N, Pines O. 2007. Eclipsed distribution: a phenomenon of dual targeting of protein and its significance. *Bioessays* **29**(8): 772-782.
Ristic Z, Bukovnik U, Momcilovic I, Fu J, Vara Prasad PV. 2008. Heat-induced accumulation of chloroplast protein synthesis elongation factor, EF-Tu, in winter wheat. *J Plant Physiol* **165**(2): 192-202.
Rudhe C, Chew O, Whelan J, Glaser E. 2002. A novel in vitro system for simultaneous import of precursor proteins into mitochondria and chloroplasts. *Plant Journal* **30**(2): 213-220.
Ruwe H, Kupsch C, Teubner M, Schmitz-Linneweber C. 2011. The RNA-recognition motif in chloroplasts. *J Plant Physiol* **168**(12): 1361-1371.
Sato N, Albrieux C, Joyard J, Douce R, Kuroiwa T. 1993. Detection and Characterization of a Plastid Envelope DNA-Binding Protein Which May Anchor Plastid Nucleoids. *Embo Journal* **12**(2): 555-561.
Sato N, Misumi O, Shinada Y, Sasaki M, Yoine M. 1997. Dynamics of localization and protein composition of plastid nucleoids in light-grown pea seedlings. *Protoplasma* **200**(3-4): 163-173.
Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **13**(12): 663-670.
Schwacke R, Fischer K, Ketelsen B, Krupinska K, Krause K. 2007. Comparative survey of plastid and mitochondrial targeting properties of transcription factors in Arabidopsis and rice. *Mol Genet Genomics* **277**(6): 631-646.
Seguí-Simarro JM, Testillano PS, Risueño MC. 2003. Hsp70 and Hsp90 change their expression and subcellular localization after microspore embryogenesis induction in Brassica napus L. *Journal of structural biology* **142**(3): 379-391.
Selldén G, Leech R. 1981. Localization of DNA in Mature and Young Wheat Chloroplasts Using the Fluorescent Probe 4'-6-Diamidino-2-phenylindole. *Plant physiology* **68**(4): 731-734.
Shen YY, Wang XF, Wu FQ, Du SY, Cao Z, Shang Y, Wang XL, Peng CC, Yu XC, Zhu SY, Fan RC, Xu YH, Zhang DP. 2006. The Mg-chelatase H subunit is an abscisic acid receptor. *Nature* **443**(7113): 823-826.
Singh BN, Mishra RN, Agarwal PK, Goswami M, Nair S, Sopory SK, Reddy MK. 2004. A pea chloroplast translation elongation factor that is regulated by abiotic factors. *Biochem Biophys Res Commun* **320**(2): 523-530.
Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**(6): 1581-1590.
Small I, Wintz H, Akashi K, Mireau H. 1998. Two birds with one stone: genes that encode products targeted to two or more compartments. *Plant molecular biology* **38**(1-2): 265-277.
Stern DB, Goldschmidt-Clermont M, Hanson MR. 2010. Chloroplast RNA Metabolism. *Annu Rev Plant Biol* **61**(1): 125-155.
Stern DB, Higgs DC, Yang JJ. 1997. Transcription and translation in chloroplasts. *Trends in plant science* **2**(8): 308-315.
Strand A, Asami T, Alonso J, Ecker JR, Chory J. 2003. Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature* **421**(6918): 79-83.
Susek RE, Ausubel FM, Chory J. 1993. Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell* **74**(5): 787-799.
Takenaka M, Verbitskiy D, van der Merwe JA, Zehrmann A, Brennicke A. 2008. The process of RNA editing in plant mitochondria. *Mitochondrion* **8**(1): 35-46.

Terasawa K, Sato N. 2005. Visualization of plastid nucleoids in situ using the PEND-GFP fusion protein. *Plant and Cell Physiology* **46**(4): 649-660.

Terasawa K, Sato N. 2009. Plastid localization of the PEND protein is mediated by a noncanonical transit peptide. *FEBS J* **276**(6): 1709-1719.

Thomas CL, Maule AJ. 2000. Limitations on the use of fused green fluorescent protein to investigate structure-function relationships for the cauliflower mosaic virus movement protein. *Journal of General Virology* **81**: 1851-1855.

Tiboni O, Di Pasquale G, Ciferri O. 1978. Purification of the elongation factors present in spinach chloroplasts. *Eur J Biochem* **92**(2): 471-477.

Tillich M, Beick S, Schmitz-Linneweber C. 2010. Chloroplast RNA-binding proteins: Repair and regulation of chloroplast transcripts. *RNA Biology* **7**(2): 172-178.

Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM. 2005. Editing of plastid RNA in Arabidopsis thaliana ecotypes. *Plant Journal* **43**(5): 708-715.

Tillich M, Hardel SL, Kupsch C, Armbruster U, Delannoy E, Gualberto JM, Lehwark P, Leister D, Small ID, Schmitz-Linneweber C. 2009. Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A* **106**(14): 6002-6007.

Vermel M, Guermann B, Delage L, Grienenberger JM, Marechal-Drouard L, Gualberto JM. 2002. A family of RRM-type RNA-binding proteins specific to plant mitochondria. *Proc Natl Acad Sci U S A* **99**(9): 5866-5871.

Vijgenboom E, Nilsson L, Bosch L. 1988. The elongation factor EF-Tu from E. coli binds to the upstream activator region of the tRNA-tufB operon. *Nucleic Acids Res* **16**(21): 10183-10197.

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat Cell Biol* **7**(12): 1224-1231.

von Braun SS, Sabetti A, Hanic-Joyce PJ, Gu J, Schleiff E, Joyce PB. 2007. Dual targeting of the tRNA nucleotidyltransferase in plants: not just the signal. *J Exp Bot* **58**(15-16): 4083-4093.

von Heijne G, Steppuhn J, Herrmann RG. 1989. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur J Biochem* **180**(3): 535-545.

Waegemann K, Soll J. 1996. Phosphorylation of the transit sequence of chloroplast precursor proteins. *J Biol Chem* **271**(11): 6545-6554.

Watanabe N, Che FS, Iwano M, Takayama S, Yoshida S, Isogai A. 2001. Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. *J Biol Chem* **276**(23): 20474-20481.

Watson JC, Surzycki SJ. 1982. Extensive sequence homology in the DNA coding for elongation factor Tu from Escherichia coli and the Chlamydomonas reinhardtii chloroplast. *Proc Natl Acad Sci U S A* **79**(7): 2264-2267.

Weinl S, Held K, Schlucking K, Steinhorst L, Kuhlgert S, Hippler M, Kudla J. 2008. A plastid protein crucial for Ca2+-regulated stomatal responses. *New Phytol* **179**(3): 675-686.

Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* **9**(5): 383-395.

Xiong JY, Lai CX, Qu Z, Yang XY, Qin XH, Liu GQ. 2009. Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKP1 to mediate transcriptional repression. *Plant Mol Biol* **71**(4-5): 437-449.

Yanagisawa S. 2001. The transcriptional activation domain of the plant-specific Dof1 factor functions in plant, animal, and yeast cells. *Plant and Cell Physiology* **42**(8): 813-822.

Yogev O, Pines O. 2011. Dual targeting of mitochondrial proteins: mechanism, regulation and function. *Biochim Biophys Acta* **1808**(3): 1012-1020.

Yoo HH, Kwon C, Chung IK. 2010. An Arabidopsis splicing RNP variant STEP1 regulates telomere length homeostasis by restricting access of nuclease and telomerase. *Mol Cells* **30**(3): 279-283.

Yoo HH, Kwon C, Lee MM, Chung IK. 2007a. Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in Arabidopsis. *Plant Journal* **49**(3): 442-451.

Yoo HH, Kwon C, Lee MM, Chung IK. 2007b. Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in Arabidopsis. *Plant J* **49**(3): 442-451.

Zhang M, Zhang B, Qian Q, Yu Y, Li R, Zhang J, Liu X, Zeng D, Li J, Zhou Y. 2010. Brittle Culm 12, a dual-targeting kinesin-4 protein, controls cell-cycle progression and wall properties in rice. *Plant Journal* **63**(2): 312-328.

VII. Publications

Paper I

Paper II

Manuscript I

Manuscript II

Manuscript III

Manuscript IV “work in progress”

ISBN xxx-xx-xxxx-xxx-x