Targeting and function of CAH1 - Characterization of a novel protein pathway to the plant cell chloroplast

STEFAN BURÉN

Akademisk avhandling

som med vederbörligt tillstånd av rektorsämbetet vid Umeå universitet för avläggande av Teknologie doktorsexamen i Växters cell- och molekylärbiologi, framläggs till offentligt försvar i KB3A9, KBC-huset, Umeå Universitet, fredagen den 29 januari 2010 klockan 10.00. Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent:
Dr. Muriel Bardor
Université de Rouen
Mont-Saint-Aignan, France

Umeå Plant Science Centre
Department of Plant Physiology
Umeå university
Sweden 2010
Targeting and function of CAH1 -
Characterization of a novel protein pathway to the plant cell chloroplast

Stefan Burén
Umeå Plant Science Centre, Department of Plant Physiology, Umeå University

ISBN: 978-91-7264-933-0

Abstract
The chloroplast is the organelle within the plant cell where photosynthesis is taking place. This organelle is originating from a cyanobacterium that was engulfed by a eukaryotic cell. As a consequence of the transition from endosymbiont to organelle, most of the cyanobacterial genes have been transferred to the host cell’s nuclear genome, resulting in the need for a massive import of gene products (proteins) back to the organelle. Until recently, this import has been believed to exclusively be mediated by a translocon complex in the chloroplast envelope (Toc-Tic), responsible for import of proteins translated in the cytosol.

We have identified a protein in the model plant Arabidopsis thaliana (CAH1) that, instead of being imported from the cytosol, is trafficking via the endomembrane system (ER/Golgi apparatus). At least part of the transport is mediated by canonical vesicle trafficking elements (from the ER to the Golgi). This novel route offers possibilities for several protein modifications, such as anchoring of asparagine (N)-linked glycans. By expression of point mutated variants of the CAH1 protein we have seen that both N-linked glycans (anchored at up to five sites on the protein), and an intra-molecular disulphide bridge, were required for correct folding, trafficking and function of the CAH1 protein. For that reason, we propose that an additional route exists as a complement to the Toc-Tic system in plants, for delivery of proteins with requirements of certain post-translational modifications. Finally, we show that CAH1 is playing a crucial role in the photosynthetic capacity of Arabidopsis. Mutant plants with disrupted CAH1 gene expression showed reduced CO2 uptake rates and accumulated less starch than wild-type plants.

Further study of the CAH1 protein is important for revealing its function in photosynthesis. Characterization of the route for CAH1 to the chloroplast might also shed some light on the evolution of the plant cell and clarify the reason for having several chloroplast import pathways working in parallel. It might also have profound effects on the possibilities of using plants as bio-factories for production of recombinant glycoproteins, which make up the vast majority of the bio-pharmaceutical molecules.

Keywords: Arabidopsis, chloroplast, endomembrane system, CAH1, protein targeting, N-glycosylation
Targeting and function of CAH1 - Characterization of a novel protein pathway to the plant cell chloroplast

STEFAN BURÉN

Umeå Plant Science Centre
Department of Plant Physiology
Umeå university
Sweden 2010
Till Peter
Abstract

The chloroplast is the organelle within a plant cell where photosynthesis takes place. This organelle originates from a cyanobacterium that was engulfed by a eukaryotic cell. During the transition from endosymbiont to organelle most of the cyanobacterial genes were transferred to the nuclear genome of the host cell, resulting in a chloroplast with a much reduced genome that requires massive import of gene products (proteins) back to the organelle. The majority of these proteins are translated in the cytosol as pre-proteins containing targeting information that directs them to a translocon complex in the chloroplast envelope, the Toc-Tic system, through which these proteins are transported.

We have identified a protein in the model plant Arabidopsis thaliana, CAH1, that is trafficked via the endomembrane system (ER/Golgi apparatus) to the chloroplast instead of using the Toc-Tic machinery. This transport is partly mediated by canonical vesicle trafficking elements involved in ER to Golgi transport, such as Sar1 and RabD GTPases. Analysis of point mutated variants of CAH1 showed that both N-linked glycans and an intra-molecular disulphide bridge are required for correct folding, trafficking and function of the protein. Since chloroplasts lack N-glycosylation machinery, we propose that a route for chloroplast proteins that require endomembrane-specific post-translational modifications for their functionality exists as a complement to the Toc-Tic system. We also show that mutant plants with disrupted CAH1 gene expression have reduced rates of CO$_2$ uptake and accumulate lower amounts of starch compared to wild-type plants, indicating an important function of the CAH1 protein for the photosynthetic capacity of Arabidopsis.

Further study of CAH1 will not only be important to reveal its role in photosynthesis, but characterization of this novel protein pathway to the chloroplast can also shed light on how the plant cell evolved and clarify the purpose of keeping several chloroplast import pathways working in parallel. In addition, knowledge about this pathway could increase the opportunities for using plants as bio-factories for production of recombinant glycoproteins, which make up the vast majority of the bio-pharmaceutical molecules.
Kloroplasten är den organell i växtcellen där fotosyntesen sker. Denna organell härrör från en cyanobakterie som togs upp av en eukaryot cell. Under omvandlingen från endosymbiont till organell har de flesta av den ursprungliga cyanobakteriens gener flyttats över till växtcellens eget kärn genom, vilket resulterat i en kloroplast som endast kan producera ett fåtal av de proteiner den behöver och som istället kräver att en mängd genprodukter (proteiner) transporteras tillbaka till organellen. De flesta av dessa proteiner syntetiseras i cytosolen som polypeptider innehållande en speciell signal för kloroplasten, och transporteras över kloroplastens dubbelmembran (envelop) med hjälp av ett specifikt importsystem (Toc-Tic).


För att kunna fastställa den exakta funktionen för CAH1 kommer ytterliga studier att vara nödvändiga. En fördjupad karakterisering av transportvägen som CAH1 följer till kloroplasten kan dessutom ge kunskap om hur växtcellen uppkom, samt besvara varför flera importvägar arbetar till synes parallellt med varandra. Kunskap om denna transportväg kan även bidra med användbar information i försöken att nyttja växter till att uttrycka rekombinanta N-glykosylerade proteiner, t. ex. antikroppar och vacciner.
List of papers

The thesis is based on the following papers, which will be referred to in the text by their Roman numerals. Paper I is printed with kind permission from Nature Publishing Group.


II. Stefan Burén, Cristina Ortega-Villasante, Göran Samuelsson, Laszlo Bako and Arsenio Villarejo: Optimization of the 2A peptide coexpression system to study trafficking of the plastid N-glycoprotein CAH1 in Arabidopsis thaliana. (manuscript)

III. Stefan Burén, Cristina Ortega-Villasante, Amaya Blanco-Rivero, Andrea Martínez-Bernardini, Tatiana Shutova, Laszlo Bako, Arsenio Villarejo, Göran Samuelsson: N-glycosylation is required for trafficking and activity of a chloroplast localized carbonic anhydrase (CAH1) in Arabidopsis thaliana. Submitted to JBC

IV. Stefan Burén*, Amaya Blanco-Rivero*, Cristina Ortega-Villasante, Göran Samuelsson, Arsenio Villarejo: Specific suppression of the chloroplast N-glycosylated carbonic anhydrase (CAH1) has major impact on the photosynthetic performance of Arabidopsis thaliana. (manuscript)

* These authors made equal contributions
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmyI-1</td>
<td>α-Amylase isoform I-1</td>
</tr>
<tr>
<td>Arf1</td>
<td>ADP-ribosylation factor 1</td>
</tr>
<tr>
<td>β-ME</td>
<td>β-mercaptoethanol</td>
</tr>
<tr>
<td>BFA</td>
<td>Brefeldin A</td>
</tr>
<tr>
<td>BiP</td>
<td>Binding Protein</td>
</tr>
<tr>
<td>CA</td>
<td>Carbonic Anhydrase</td>
</tr>
<tr>
<td>CAH1</td>
<td>Carbonic Anhydrase 1 in <em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>CHX</td>
<td>Cycloheximide</td>
</tr>
<tr>
<td>CNX</td>
<td>Calnexin</td>
</tr>
<tr>
<td>ConA</td>
<td>Concanavalin A</td>
</tr>
<tr>
<td>CRT</td>
<td>Calreticulin</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic Reticulum</td>
</tr>
<tr>
<td>ERES</td>
<td>ER-Export Site</td>
</tr>
<tr>
<td>ERQC</td>
<td>ER-Quality Control</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed Sequence Tag</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-Activated Cell Sorting</td>
</tr>
<tr>
<td>GFP/CFP/YFP</td>
<td>Green/Cyan/Yellow Fluorescent Protein</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-Acetylglucosamine</td>
</tr>
<tr>
<td>GNT I</td>
<td>N-Acetylglucosaminyl Transferase I</td>
</tr>
<tr>
<td>GT</td>
<td>UDP-glucose:glycoprotein Glucosyl Transferase</td>
</tr>
<tr>
<td>GUS</td>
<td>β-Glucuronidase</td>
</tr>
<tr>
<td>HA-CAH1</td>
<td>HA-tagged wild type CAH1</td>
</tr>
<tr>
<td>IP</td>
<td>Immunoprecipitation</td>
</tr>
<tr>
<td>KDEL</td>
<td>ER retention signal</td>
</tr>
<tr>
<td>NPP1</td>
<td>Nucleotide Pyrophosphate/Phosphodiesterase 1</td>
</tr>
<tr>
<td>OST</td>
<td>Oligosaccharyl Transferase</td>
</tr>
<tr>
<td>PLAM</td>
<td>Plastid Associated Membrane</td>
</tr>
<tr>
<td>PNGase F</td>
<td>Peptide-N-Glycosidase F</td>
</tr>
<tr>
<td>RabD2a</td>
<td>Member of plant D subclass of the Rab family of small GTPases</td>
</tr>
<tr>
<td>Rubisco</td>
<td>Ribulose-1,5-bisphosphate carboxylase oxygenase</td>
</tr>
<tr>
<td>Sar1</td>
<td>Secretion associated, ras-related protein1 (small GTPase)</td>
</tr>
<tr>
<td>SP</td>
<td>Signal Peptide for the ER</td>
</tr>
<tr>
<td>SPP</td>
<td>Stromal Processing Peptidase</td>
</tr>
<tr>
<td>SRP</td>
<td>Signal Recognition Particle</td>
</tr>
<tr>
<td>Tat</td>
<td>Twin-arginine tranlocation</td>
</tr>
<tr>
<td>Toc-Tic</td>
<td>Translocon of outer/inner chloroplast envelope membrane</td>
</tr>
<tr>
<td>TP</td>
<td>Transit Peptide</td>
</tr>
<tr>
<td>wt</td>
<td>wild type</td>
</tr>
</tbody>
</table>
Contents

List of papers ........................................... ix
Abbreviations ........................................... x

Contents xi

1 Background 3
1.1 Central dogma ........................................ 3
1.2 Chloroplast ........................................... 4
  1.2.1 Photosynthesis ................................... 5
  1.2.2 Chloroplast evolution ............................... 6
  1.2.3 Protein import into chloroplasts ................. 8
1.3 Endomembrane system ................................ 10
  1.3.1 Endoplasmic reticulum ............................ 10
  1.3.2 Golgi apparatus ................................... 11
  1.3.3 Protein trafficking between the ER and the Golgi apparatus ............... 11
  1.3.4 Protein folding and post-translational modification in the ER and Golgi ........... 13
1.4 Carbonic anhydrases .................................. 17
  1.4.1 The .................................................. 18

2 Results achieved 19
2.1 Paper I ............................................... 19
  2.1.1 CAH1 is lacking N-terminal transit peptide for targeting to the chloroplast .............. 20
  2.1.2 CAH1 contains an N-terminal signal peptide for the ER, but is localized to the chloroplast .............. 23
  2.1.3 Chloroplast localized CAH1 is N-glycosylated in the ER 23
  2.1.4 Attachment of complex type glycans to CAH1 suggest a route to the chloroplast via the Golgi apparatus .......... 25
  2.1.5 Trafficking of CAH1 is blocked by Brefeldin A .............. 26
2.2 Manuscript II .......................................... 26
  2.2.1 Development of an optimized 2A peptide co-expression system ...................... 28
  2.2.2 Released GTPase proteins were functional .............. 30
  2.2.3 Endo H resistance assays can reveal information about trafficking ...................... 31
  2.2.4 Potential areas of application for the 2A peptide co-expression system ...................... 32
2.3 Manuscript III .......................................... 34
  2.3.1 CAH1 harbours four or five N-glycans ...................... 35
  2.3.2 A disulphide bridge is important for folding and ER-export of CAH1 ...................... 36
  2.3.3 Non-glycosylated CAH1 forms aggregates ...................... 37
  2.3.4 Evolutionary aspects of CAH1 ...................... 37
2.4 Manuscript IV ......................................................... 38
  2.4.1 Starch content and CO₂ exchange rates are reduced in
        CAH1 mutants ............................................... 39

3 Conclusions ......................................................... 41
  3.1 Future perspectives ........................................... 42

4 Acknowledgement ................................................ 45

5 References ......................................................... 47
A well-known fact, emphasized in many recent dissertations, is that plants cannot move. At least not very fast (Sjödin, 2007). Another important feature in which plants differ from animals is their ability to fix inorganic carbon into organic molecules. This process, photosynthesis, takes place in the chloroplast, which is an organelle within the plant cell. In addition to converting otherwise inaccessible carbon dioxide into sugars, photosynthesis splits water to get the reducing power needed for carbon fixation. As a by-product from this reaction molecular oxygen is released into the atmosphere and subsequently used by you and me when breathing. In addition to the apparent reasons to study this remarkable process, photosynthetic research has gained an increased interest the last years because of the ambition to exchange an oil-based economy for a society where renewable energy resources are used.

The aim of this thesis has not been photosynthesis specifically, but rather the study of a protein located within the photosynthetic organelle, the chloroplast. Although some recent results have indicated that this protein is involved in or connected to photosynthesis or photosynthetic processes, the focus of the results presented here concern aspects that make this protein unique, or at least different, of most other chloroplast proteins, e.g. its intracellular transport and biochemical properties.

While many recent dissertations have dealt with large scale genomic, proteomic or metabolomic data analysis, this thesis focuses on one gene/protein, with the ambition to draw general conclusions from the obtained information.
Chapter 1

Background

The chloroplast is an organelle within the plant cell. It is surrounded by a double membrane and possesses its own genome. Although most of the genes encoding chloroplast proteins have been transferred to the nucleus of the plant cell, analysis of the chloroplast genome has made it clear that the chloroplast originated from a cyanobacterium that was engulfed by a eukaryotic cell. Since the majority of the cyanobacterial genes have been relocated to the nucleus, massive transport of proteins back to the chloroplast must take place. Until recently, all proteins were believed to cross the chloroplast envelope through a protein complex, capable of translocating unfolded proteins from the cytosol to the chloroplast interior. This work describes a protein which is instead trafficked through the endomembrane system to the chloroplast stroma. The results raises, rather than answers, new questions about chloroplast function, regulation and evolution.

1.1 Central dogma

As the name implies, the so-called central dogma is key to the understanding of biology in general and of molecular biology in particular. It describes the normal flow of biological information in the cell, where the first step is transfer, or copy, of the genetic information from DNA to RNA in a process termed transcription. The RNA is then translated into a polypeptide, made up from amino acids and joined together by peptide bonds in a specific order which is determined by the RNA code. This chain of amino acids is then folded into a protein with a higher-level of structure responsible for the activity and function of the protein.

Transcription, or synthesis of RNA, takes place in the nucleus (for nuclear encoded proteins). The RNA (called mRNA, for messenger RNA) is then transported out of the nucleus and translated by free or membrane bound ribosomes depending of the final destination of the synthesized protein (Figure 1.1).

While the central dogma was first articulated in 1958, protein modifications have during recent years attracted increased attention for their importance in altering the basic properties of the protein given by the genetic code. Protein
modifications can happen during or after synthesis of the polypeptide. Examples of common modifications are phosphorylation, glycosylation, oxidation and acetylation. Modified sites in proteins not only change their individual functions, but can also work together in order to fine-tune molecular interactions and stability, activity, localization, targeting and folding (Jensen, 2006).

### 1.2 Chloroplast

Fundamental for the understanding of the results presented in this thesis is a basic knowledge of the organization and the function of the chloroplast. The chloroplast is a type of plastid, a subcellular organelle found in cells of plants and algae, capable of performing photosynthesis. It is separated from the surrounding cytoplasm of the eukaryotic plant cell by a double membrane, called the chloroplast envelope. The chloroplast has six distinct suborganellar compartments: three different membranes (inner and outer envelope membranes together with thylakoid membranes), an intermembrane space between the two envelope membranes, a soluble interior between the inner membrane and the thylakoid membranes called the stroma, and an aqueous lumen within the thylakoids (Jarvis, 2008) (Figure 1.2). The chloroplast has its own functional genome, although most of the genes have been transferred to the nucleus of the cell. The main function of the chloroplast is to perform photosynthesis, but the chloroplast is also the site of fatty acid biosynthesis, nitrate assimilation and amino-acid biosynthesis (Waters and Langdale, 2009).
1.2. CHLOROPLAST

Figure 1.2. The plant cell chloroplast. The subcellular organization of the chloroplast includes three different membrane systems: the outer and inner envelope membranes and the thylakoid membrane. There are three additional compartments: the inter-membrane space (between the outer and the inner envelope membranes), the soluble stroma and finally the aqueous lumen within the thylakoid membranes.

1.2.1 Photosynthesis

The first organisms to carry out oxygenic photosynthesis were the cyanobacteria (Björn and Govindjee, 2009). Or to put this in other words; at a certain point in history, a photosynthetic machinery was assembled in which sunlight was used as a power source to perform the thermodynamically and chemically demanding reaction of splitting water in order to get reducing equivalents needed to convert carbon dioxide into sugars (and then into other organic molecules). The organism where this event took place was a cyanobacterium. Following this, the almost unlimited resources of water on earth could be exploited and the prerequisites for life substantially changed. In addition and as a by-product from this reaction molecular oxygen was released, increasing the metabolic efficiency due to the possibility of aerobic respiration. Hence, the evolution of the oxygenic photosynthetic process offered an advantage for organisms capable of undergoing aerobic respiration, relegating the majority of anaerobic organisms to isolated environments. Additionally, oxygen evolution enabled a protective ozone layer against the harmful UV radiation to be formed (Barber, 2008). Several innovations were required for this to take place: the use of two photosystems working in a series capable of generating enough difference in redox potential for water oxidation and reduction of carbon dioxide, an oxygen evolving complex, an "electron buffer" (a plastoquinonone pool) and modifications of the pigments attached to the reaction centres (Björn and Govindjee, 2009).

Photosynthesis in plant cells takes place inside the chloroplast. The photosynthetic process is usually divided in two distinct phases, one consisting of
the light reactions in the thylakoid membranes in which the light-driven flow of electrons through the multi-subunit complexes mentioned above results in chemical energy (ATP) and reducing power (NADPH), and the second consisting of the carbon-fixing reactions (light-independent reactions) in the stroma where the ATP and NADPH are used by the Rubisco and other enzymes of the Calvin cycle to fix carbon dioxide and to generate sugars (Waters and Langdale, 2009). These sugars can then be metabolized into other carbohydrates and used for the production of energy, amino acids and lipids.

1.2.2 Chloroplast evolution

Although first met with scepticism, the hypothesis that chloroplasts are derived from cyanobacteria (at that time known as blue-green algae), proposed by Mereschkowsky in 1905, was later supported by electron microscopy and biochemical studies (Raven and Allen, 2003). Now genomic analyses have concluded that the chloroplast indeed originates from a secondary endosymbiotic event, in which a eukaryotic cell already possessing mitochondria (as a result of a primary endosymbiotic event between an archaea and a proteobacterium in which the latter gave rise to mitochondria), engulfed a cyanobacterium (Kilian and Kroth, 2003; Kilian and Kroth, 2005; McFadden, 2001; Raven and Allen, 2003; Reyes-Prieto et al., 2007) (Figure 1.3).

![Endosymbiotic events](image)

**Figure 1.3. Endosymbiotic events.** The eukaryotic cell evolved from a primary endosymbiotic event between an archaea with its own genome (orange) and a proteobacterium (brown), in which the latter gave rise to the mitochondria. This organism later engulfed a cyanobacterium (green) that became the origin for the chloroplast (secondary endosymbiotic event). In some cases additional endosymbiosis occurred, giving rise to several different organisms, including numerous algal phyla (brown algae, etc) as well as apicomplexans, such as Plasmodium (modified from Raven and Allen, 2003).

Many of the proteins required for function of the chloroplast are today encoded by the nuclear genome as a result of a process called gene transfer. Studies
of plastid genomes imply that chloroplasts only possess genes for about 60-200 proteins, depending on the organism (Leister, 2003; Martin et al., 2002), while as many as 5000 proteins might be targeted to the chloroplast.

Comparison of the *Arabidopsis thaliana* genome with several cyanobacteria, yeast and prokaryote genomes suggest that almost one fifth (4500 genes) of the total Arabidopsis genes originate from the cyanobacterial ancestor of the chloroplast (Martin et al., 2002). In other studies, between 1400 to 1500 (Abdallah et al., 2000) or 400 to 2200 (Rujan and Martin, 2001) Arabidopsis genes of cyanobacterial origin have been proposed. However, gene origin and protein compartmentalization do not strictly correspond since most of these gene products with a cyanobacterial origin are not targeted back to the chloroplast, suggesting a massive redistribution of cyanobacterium-derived proteins to other cellular compartments, or they have been lost as the endosymbiont evolved. On the other hand, many non-cyanobacterial proteins are imported to the chloroplasts (Leister, 2003).

Interestingly, and discussed later on, many of the proteins of cyanobacterial origin are predicted to enter the secretory pathway (Martin et al., 2002). Obviously, selective advantages for transfer of most genes to the host nucleus exist. Whether this is an ongoing process that will end up with an organelle without its own genome only time can tell. The majority of genes that still remain in the chloroplast genome seem to be involved in photosynthesis or transcription and translation of the chloroplast genes, while most other genes are now encoded in the nucleus (Reyes-Prieto et al., 2007). This suggest that there is an evolutionary reason why some genes have been kept within the organelle while others not.

Establishment of the chloroplast from the endosymbiont involved more than arranging a functional system for import of proteins whose genes were transferred to the nucleus. Intuitively it would seem that engulfment of a cyanobacterium with two membranes would have resulted in an organelle with three membranes, where the inner two membranes originate from the prokaryote and the outer one originates from the outer membrane of the phagotrophic organism. This is in contrast to the two membranes of the chloroplast envelope of today. Cellular membranes are very well conserved and characterisation of the lipid and protein content has been performed in order to answer this question. Interestingly, biochemical analyses of the membranes are ambivalent. While the inner envelope membrane very likely correspond to the plasma membrane of the cyanobacterial ancestor, the outer envelope membrane has both components found in cyanobacterial outer membranes (high content of galactolipids and carotenoids) and elements pointing to an eukaryotic origin (phosphatidylcholine) (Kilian and Kroth, 2003). This suggests that the outer envelope membrane of chloroplasts today is a chimera resulting from a fusion of the two outer membranes of the early endosymbiont and perhaps supporting this idea, the components of the Toc-Tic complex are a mix of endosymbiotic and eukaryotic origin (Bhattacharya et al., 2007).
1.2.3  Protein import into chloroplasts

Most chloroplast proteins are encoded in the nucleus. As a consequence of this genomic re-organization, most chloroplast gene transcripts are translated into polypeptides by cytosolic ribosomes and require further targeting of the proteins "back" to the chloroplast (Jarvis, 2008). Until recently, all proteins destined for the chloroplast interior were thought to possess an amino terminal extension, called transit peptide (TP), directing the pre-protein to a translocon complex in the chloroplast envelope. Although recent proteomic studies have identified some exceptions, in which chloroplast protein precursors lack this type of targeting signal (Kleffmann et al., 2004), the vast majority of the chloroplast proteins are believed to be directed to the chloroplast by a mechanism based on a TP pre-sequence and this pathway is considered as the canonical chloroplast protein import system (Cline and Dabney-Smith, 2008).

The TPs of these precursor proteins function as 'zip codes' or signal sequences that are recognized by cytosolic chaperones, whose binding prevent folding of the polypeptides and directs them to the translocon system in the chloroplast envelope. This import system, called the Toc-Tic complex (translocon at the outer and inner envelope membranes of chloroplasts), is formed by two multi-subunit complexes (Toc in the outer chloroplast membrane and Tic in the inner, respectively) that together enable post-translational translocation across the two envelope membranes as well as regulating the import process itself (Bedard and Jarvis, 2005). In addition, multiple homologues of the Toc receptor components are found which might be involved in recognition and controlling import specificity. Recently, it has also been reported that the Toc-Tic complex is under redox regulation (Stengel et al., 2009).

In short, the TP of the pre-protein to be imported is recognized by receptors at the chloroplast surface (TOC159 and TOC34). These components of the Toc have been shown to possess GTPase activity (Cline and Dabney-Smith, 2008; Jarvis, 2008). The third component of the Toc-core, TOC75, is embedded in the outer membrane and forms a pore due to its β-barrel structure and transports the largely unfolded pre-protein across outer membrane in a GTP-dependent process (Jarvis and Robinson, 2004). Once at the inter-membrane space, HSP70 family chaperones, TOC12 and TIC22 mediate the interaction between Toc and Tic. The extended polypeptide is subsequently translocated across the inner membrane through a Tic protein-conducting channel involving TIC20 and TIC110. This step requires high levels of ATP in the stroma and the presence of stroma located molecular chaperones (Bedard and Jarvis, 2005). Although the precise mechanism by which the Toc complex functions remains to be clarified, even less is known about the Tic complex. Soon upon arrival in the stroma, the TP is cleaved off by a stromal processing peptidase (SPP) and degraded (Jarvis, 2008), resulting in a mature stromal protein or revealing additional targeting information responsible for further directing of the polypeptide to the thylakoid (Bhattacharya et al., 2007) (Figure 1.4).
Figure 1.4. Chloroplast protein import by the Toc-Tic complex. The pre-protein, surrounded by cytosolic chaperones, interacts with receptor components of the Toc complex and is transferred across the outer envelope membrane. The complex contacts with components of the Tic apparatus at the inner membrane and the pre-protein is translocated simultaneously across both envelope membranes. Afterwards, the TP is removed by stromal processing peptidase (SPP) releasing the mature protein into the stroma (modified from Jarvis 2008).

The TP of chloroplast targeted proteins are remarkably heterogenic, ranging from 20 to > 100 amino acid residues. The only conserved properties seem to be an abundance of hydroxylated residues (serine in particular) and an overall positive charge (Jarvis, 2008; Jarvis and Robinson, 2004). In addition, they appear not to form any secondary structure but instead have a random coil conformation, which might explain the recruitment of cytosolic factors. Another possibility is that structure formation requires binding of TP to specific lipids at the outer envelope membrane. Since translocation of the pre-protein through the Toc-Tic complexes requires the polypeptide to assume an extended conformation, cytosolic chaperones are thought to assist in delivering the unfolded polypeptide to the chloroplast surface. This suggests that the TP is designed to attract binding of such cytosolic factors (Jarvis, 2008; Jarvis and Soll, 2002).

Although most chloroplast proteins follow the post-translational pathway through the Toc-Tic complex, some exceptions have recently been described. Firstly, most outer envelope proteins are inserted in the membrane from the cytosolic side without cleavable transit peptide but directed by intrinsic targeting information, with or without help from Toc components (Jarvis, 2008).
Also, two proteins (ceQORH and Tic32) have been shown to be targeted to the inner membrane without cleavable targeting signals (Miras et al., 2002; Nada and Soll, 2004). In addition and as later presented in this thesis, there seems to be a pathway for chloroplast proteins to be transported via the endomembrane system.

1.3 Endomembrane system

The plant endomembrane contains several membrane bound organelles (Figure 1.5). While endoplasmic reticulum (ER) and the Golgi apparatus will be presented in more detail later, plant cells also contain at least two types of endosomes: early endosomes involved in sorting and recycling and late endosomes/prevacuolar compartments en route to the lytic vacuole. Two major functional types of vacuoles also exist in plant cells: the lytic vacuole and the protein-storage vacuoles. Lytic vacuoles function as compartments for degradation and waste storage, while protein-storage vacuoles accumulate proteins mainly used as nutrients during seed germination. These two types of vacuoles can fuse, giving rise to a large central vacuole. During cell division another transient compartment is also formed, the cell plate, by fusion of transport vesicles (Jurgens, 2004).

The default secretory pathway leads from the ER via the Golgi apparatus to the plasma membrane. Proteins not destined for the plasma membrane are sorted in the Golgi. Proteins aimed to remain in the ER, such as ER located chaperones, are recognized by the presence of well-characterized ER retention signals (H/KDEL). Other signals are responsible for sorting of membrane proteins or proteins destined for vacuoles.

1.3.1 Endoplasmic reticulum

The ER is a membrane-bound tubular network that stretches out from the nuclear envelope towards the plasma membrane (Jurgens, 2004). The ER network of higher plants overlies the actin cytoskeleton rather than microtubules as in animal cells (Runions et al., 2006; Sparkes et al., 2009; Staehelin, 1997). Traditionally the ER was classified as smooth or rough depending on the absence or presence of membrane-bound ribosomes, however now two morphological forms are used, cisternal and tubular, which better reflect the highly dynamic nature of this organelle (Sparkes et al., 2009). Several cellular functions are allotted to the ER, such as biosynthesis of phospholipids and synthesis, post-translational modification, folding, quality-control of secreted proteins and glycoproteins, as well as regulating cytosolic calcium levels. The multifunctional nature of the ER is also exemplified by the vast number of sub-regions or domains of which the ER is composed (Staehelin, 1997).
One of the two sub-domains of the ER important for the work presented in this thesis is the rough ER, where membrane-bound ribosomes are attached to the ER. This is the entry point for proteins into the so-called secretory pathway. Another important domain of the ER, and perhaps the most dynamic and controversial, is the ER export site (ERES). This is believed to be the site for transport of soluble and membrane proteins and lipid cargo from the ER to the Golgi. Although the exact nature of the interface between the ER and the Golgi remains to be solved, some speculate whether the Golgi itself can be considered as a specialized domain of ER and that the ERES has the possibility of initiating the biogenesis of a new Golgi stack (Sparkes et al., 2009). Transport between the ER and the Golgi apparatus is bidirectional and believed to be mediated by different coated vesicles. COPII vesicles are thought to function in the anterograde transport from ER to Golgi, while COPI vesicles are working in retrograde Golgi-to-ER transport (Hawes et al., 2008).

### 1.3.2 Golgi apparatus

The Golgi apparatus in plant cells is organized as individual stacks, containing several morphologically distinct cisternae from cis to trans, followed by a trans-Golgi network (Figure 1.5) (Jurgens, 2004). Golgi stacks are highly mobile and close association and/or direct contact with the plant ER has been reported in electron microscopy studies (Brandizzi et al., 2002). Also, using photo-activated GFP, Golgi bodies were seen to move with the same rate and in the same direction as ER, demonstrating that Golgi is moving with, and not over the ER (Runions et al., 2006), which supports the idea that there is a tight connection between ERES and the Golgi bodies. In recent studies using laser-trapping technology, it was shown that capture and manipulation of individual Golgi bodies in cells with depolymerised actin cytoskeleton not only resulted in movement of the Golgi stack, but also in extension or growth of the associated ER tubule, supporting the theory that Golgi bodies can possess an attachment to the ER (Sparkes et al., 2009).

The Golgi apparatus is not only functioning as a sorting station for cargo delivery to different destinations, but also as a specialised protein modifying factory where several enzymes, among them those responsible for modifying the N-linked glycans on glycoproteins passing the Golgi on the way to their final destination (Jurgens, 2004), reside in different subdivisions of the Golgi stacks (cis-, medial- and trans-cisternae) (Figure 1.5).

### 1.3.3 Protein trafficking between the ER and the Golgi apparatus

Protein synthesis is carried out by ribosomes that translate the genetic information from the RNA molecule into a polypeptide with a specific sequence
of amino acids. These ribosomes are either free in the cytosol (or inside the endosymbiotic organelles mitochondria and chloroplasts) or bound to the cytosolic face of the ER membrane. The free ribosomes are synthesizing cytosolic, nuclear, peroxisomal and most of the mitochondria and chloroplast proteins, while proteins destined for compartments within the endomembrane system (ER, Golgi, endosomes, vacuole), the plasma membrane or secreted are synthesized by ER associated ribosomes (Hebert and Molinari, 2007) (Figure 1.5).

Ribosomes translating proteins destined for the ER must be brought in close proximity to the ER membrane in order for the nascent polypeptide to be able to cross the ER membrane. This is accomplished by the presence of a signal sequence at the N-terminus of the emerging polypeptide. This so called signal peptide (SP) contains hydrophobic amino acids and precedes the mature proteins. Appearance of this hydrophobic peptide is recognized by a multisubunit complex, called the signal recognition particle (SRP), which binds to the signal sequence and pauses translation. Protein synthesis is only resumed when the complex has come to contact with the ER and the ribosome is localized at a proteinous channel in the ER membrane. In addition to bind to SRP, the hydrophobic SP facilitates co-translational insertion of the polypeptide into the ER lumen. Upon entry in the ER, the SP is usually cleaved off from the mature protein (Hebert and Molinari, 2007). The first step in protein trafficking along the default secretory pathway is vesicle transport from the ER to the Golgi (Figure 1.5). Although very little evidence for the existence of COPII vesicles exists, they are believed to mediate anterograde traffic between the ER and the Golgi (Faso et al., 2009; Hawes et al., 2008). Formation of the COPII vesicles requires the GTPase Sar1 and its GDP/GTP exchange factor Sec12. At the cis-Golgi side, another GTPase (RabD2a) is involved in fusion of the vesicles (Hawes et al., 2008; Jurgens, 2004). Dominant mutant versions of both proteins, where the GTPase activity has been disrupted, have been shown to block trafficking of soluble proteins in the ER (Batoko et al., 2000; Takeuchi et al., 2000).

Retrograde traffic from the Golgi to the ER depends on COPI vesicles and the action of the ADP-ribosylation factor 1 (Arf1) GTPase. In contrast to the COPII-mediated bulk flow mechanism for ER exit (Hanton et al., 2006), transport of soluble and membrane proteins from the Golgi to the ER require targeting signals (Matheson et al., 2006). In a similar way as for Sar1 and RabD2a, a dominant-negative version of Arf1 was shown to inhibit Golgi-to-ER traffic (Takeuchi et al., 2002). Interestingly, inhibition of COPI function also resulted in impaired ER and disrupted anterograde transport, emphasizing that a balance between the two systems is required for normal ER and Golgi organization (Faso et al., 2009; Stefano et al., 2006).
1.3.4 Protein folding and post-translational modification in the ER and Golgi

The lumen of the ER has a more oxidizing milieu than the cytosol and contains a myriad of enzymes that are able to perform modifications on the newly synthesized proteins. In addition, many ER-resident proteins prevent protein aggregation and maintain the emerging polypeptides in a state that allows co- and post-translational modifications to take place (Hebert and Molinari, 2007; Vitale and Ceriotti, 2004). Of particular importance and interest for the work presented in this thesis are the formation of intra- and inter-molecular disulphide bonds between cysteine residues and the anchoring of N-linked oligosaccharides to the polypeptide backbone. These modifications are important for proper structural maturation of many proteins and can also be
sensed by a quality control system present in the ER (ERQC) in order to ensure that only correctly folded proteins are trafficking out of the ER, while misfolded proteins are either refolded to achieve correct structure or being sent for degradation (Crofts et al., 1998) (Figure 1.6).

Figure 1.6. The ER N-glycoprotein "quality control". Nascent polypeptide chains enter the ER lumen and N-glycans containing N-acetylglycosamine, mannose and glucose molecules are attached to asparagines residues in a site-dependent manner. The two terminal glucoses of the glycan are rapidly trimmed by sequential action of the glucosidase I and II. Mono-glucosylated N-glycans mediate initial association of folding polypeptides with the ER lectin-chaperones calnexin (CNX) and/or calreticulin (CRT) and undergo exposure to glycoprotein oxidoreductases, releasing the properly folded protein, which is rapidly deglucosylated, partially demannosylated and eventually leaves the ER. Proteins not completely folded are kept in the CNX and/or CRT cycle: The folding intermediate is released from the lectin chaperones and deglucosylated. Forward transport is inhibited by a glucosyl transferase, adding a glucose residue to glycoproteins with nearly native conformation, which undergo additional folding attempts. Released glycopolypeptides displaying major folding defects attract BiP and are extensively demannosylated and dislocated across the ER membrane for proteasome mediated degradation (modified from Hebert and Molinari, 2007).

One of the best studied chaperones in the ER is the binding protein (BiP). Unless N-glycans are added to the very N-terminus of the emerging polypeptide, BiP is the first chaperone the emerging polypeptide faces upon arrival in the ER lumen. BiP counteracts misfolding by binding to hydrophobic domains of the nascent protein, thereby preventing hydrophobic regions of different polypeptides from aggregating. When folding is complete, hydrophobic regions are no longer exposed and the protein is released from BiP (Pimpl
et al., 2006). In a similar way, BiP also prevents formation of non-native disulphide bonds that otherwise could result in protein aggregates. Another function of BiP is to bind to severely and permanently misfolded proteins and to assist in translocation of these proteins out of the ER for degradation in the cytosol (Hebert and Molinari, 2007). Recently, a role for BiP in transport of misfolded proteins to lytic vacuoles has also been proposed (Pimpl et al., 2006).

Many of the proteins entering the secretory pathway are N-glycosylated. Attachment of one or several N-linked glycans not only changes the properties of the proteins, but also assists in folding due to the action of carbohydrate-binding chaperones, even called lectin chaperones. Although much of our current knowledge about ERQC and glycan processing comes from studies on yeast and mammalian systems, most components seem to be conserved in plants (Hong et al., 2008; Parodi, 2000). While BiP binds to hydrophobic regions of the polypeptide backbone, the lectin chaperones bind to the bulky hydrophilic oligosaccharide groups (Hebert and Molinari, 2007).

The first step in the maturation of N-linked glycoproteins in the secretory pathway is the transfer of an oligosaccharide precursor (Glc3Man9GlcNAc2) from a dolichol lipid carrier to a specific asparagine residue (Asn-X-Ser/Thr, where X is any residue except Pro) on the emerging polypeptide by the oligosaccharyl transferase (OST) multisubunit complex (Lerouge et al., 1998). This precursor is subsequently modified by glucosidases and glucosyl transferases along the secretory pathway. The first modification of the N-glycan precursor that takes place in the ER is the trimming of the three glucose units (Figure 1.6). The outermost \(\alpha_1,2\)-glucose unit is hydrolyzed by glucosidase I, while the following two \(\alpha_1,3\)-linked glucose units are removed by glucosidase II (Crofts et al., 1998; Leonard et al., 2009). Trimming of the first \(\alpha_1,3\)-glucose results in monoglucosylated N-glycans (GlcMan9GlcNAc2) that are recognized by the lectin chaperones calnexin (CNX) and calreticulin (CRT). This causes the folding process to slow down and increases the efficiency in the formation of correct disulphide bonds by oxidoreductases (Hebert and Molinari, 2007). When folding is complete, the protein is released and the last glucose is removed by glucosidase II, rendering an oligosaccharide structure known as high-mannose type N-glycan. In mammals, an ER-localized mannosidase has been shown to remove one mannose residue of the correctly folded protein to yield Man8GlcNAc2, before ER-export and trafficking of the glycoprotein to the Golgi. Such mannosidase has not yet been found in plants, although ER-resident glycoproteins with that exact structure have been identified, suggesting that a similar mannosidase also exists in plant cells (Navazio et al., 1996).

Glycoproteins that have not acquired the correct folding are retained in the ER in one of two ways. If misfolding is severe, or if repeated cycles of attempted folding of the protein fail, the protein attracts binding of BiP which forms aggregates with the misfolded protein, protecting the ER from exposure
to such potentially harmful molecules and assisting in translocation of the protein out of the ER for degradation (Hebert and Molinari, 2007; Hong et al., 2008; Li et al., 2009; Parodi, 2000). If the protein is only partially misfolded, transient reglucosylation by the luminal enzyme UDP-glucose:glycoprotein glucosyl transferase (GT) again results in a monoglucosylated protein that can be recognized by the lectin chaperones CNX and CRT. These chaperones will then remain and assist in refolding of the protein (Hammond et al., 1994; Jin et al., 2009; Jin et al., 2007; Soussilane et al., 2009). Binding of CNX and CRT exposes the polypeptide to oxidoreductases that assist in formation of disulphide bonds and isomerases capable of rearranging non-native bonds. This cycling of glucosidase II and GT activities drives binding and release to the CNX/CRT chaperones and assists in proper folding of the glycoprotein until correct structure is achieved and the protein is structurally competent for export out of the ER. An alternative outcome is that folding is not successful and the polypeptide is instead sent for degradation. N-linked glycosylation therefore plays an important role in the ERQC system and indicates that folding can depend on an ensemble of different protein modifications that all must work together for a properly folded protein to appear.

Upon arrival at the Golgi apparatus, plant N-glycans can be further modified into complex-type N-glycans during the transport of the glycoprotein from cis, through medial to trans cisternae of the Golgi. First, the α-mannosidase I removes one to four α1,2-mannose residues, resulting in Man5GlcNAc2 (Figure 1.7). Then N-acetylglucosaminyl transferase I (GNT I) transfers an N-acetylglucosamine (GlcNAc) residue to the α1,3-mannoside branch of Man5GlcNAc2, to yield GlcNAcMan5GlcNAc2. Two additional mannoses are then removed by α-mannosidase II and another GlcNAc is transferred to the α1,6-mannoside branch by GNT II, resulting in GlcNAc2Man3GlcNAc2. Further action of Golgi localized glucosyl transferases results in plant specific N-glycans. Transfer of β(1,2)-xylose to the β-mannose and a(1,3)-fucose to the proximal GlcNAc core seem to be independent events occurring in the medial and trans cisternae of the Golgi. Later, additional modification of the complex-type glycan by transfer of fucose and galactose residues by β1,3-galactosyl transferase and α1,4-fucosyl transferase to the terminal GlcNAc residues might take place, resulting in antennae with Galβ1-3(Fucα1-4)GlcNAc sequences. These structures are also known as Lewis a antigens and found on the cell surface of mammalian cells and involved in cell-cell recognition and cell adhesion processes (Lerouge et al., 1998; Rayon et al., 1998). Additional modification of the N-linked glycan can take place during the transport to or in the final destination, such as for vacuoles (Lerouge et al., 1998; Rayon et al., 1998).

The reason for the occurrence of such plant-specific N-glycans is not known. A mutant allele of Arabidopsis, defective in GNT I and unable to produce complex-type N-glycans, did not show any obvious phenotype when grown under standard conditions (von Schaewen et al., 1993). However, mutation of this enzyme in mammalian cells are deleterious (Ioffe and Stanley, 1994).
1.4 Carbonic anhydrases

Carbonic anhydrase (CA) is a ubiquitous zinc-containing metalloenzyme that catalyzes the reversible hydration of CO$_2$ (Khalifah, 1971). The enzyme was first discovered in red blood cells but has since then been found in most organisms, including animals, plants, algae and some bacteria (Hewett-Emmett and Tashian, 1996). CA is important in many physiological functions that involve carboxylation and decarboxylation reactions, including both photosynthesis and respiration. CA also participates in pH regulation, inorganic carbon transport, ion transport, water and electrolyte balance (Badger and Price, 1994). The known CAs can be grouped into four distinct classes on basis of their amino acid sequence: \( \alpha \), \( \beta \), \( \gamma \) and \( \delta \). These classes have no primary sequence similarities and they are assumed to have evolved independently. The animal CAs belongs to the \( \alpha \)-family, while other eukaryotes encode \( \alpha \), \( \beta \) and \( \delta \) classes of CA (Moroney et al., 2001). Arabidopsis contains genes encoding \( \alpha \)-, \( \beta \)- and \( \gamma \)-CAs (Fabre et al., 2007). The \( \beta \)-type is the prevailing class with genes being targeted to different sub-cellular compartments such as the chloroplasts, mitochondria, plasma membrane and the cytosol. The \( \gamma \)-family encodes five genes that are targeted to the mitochondria (Harvey Millar et al., 2001). The \( \delta \)-type CA has so far only been identified in diatoms.
The requirement of a CA activity for photosynthesis, as well as for any biological system, is obvious. In the plant chloroplast, the non-catalyzed interconversion of CO$_2$ and HCO$_3^-$ is considered to be 104 times slower than the biological flux needed for CO$_2$ fixation by Rubisco (Badger and Price, 1994). If interconversion between these two species is important for the supply of CO$_2$ to the active site of Rubisco, then CA activity would be required to enable effective photosynthesis in the chloroplast stroma. Despite this important function, little progress has been made in fully elucidating the role of CAs in C3 photosynthesis.

1.4.1 The α-CAs in Arabidopsis

At least eight genes encoding α-type CAs are present in Arabidopsis thaliana (AtαCA1-8). Although the essential amino acids are present in the predicted gene products from all eight CA genes (Fabre et al., 2007), suggesting that they are functional isozymes, expressed sequence tags (ESTs) have only been reported to The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) for five of them (CA1, CA2, CA3, CA5 and CA8, as of November 2009) This indicates that CA4, CA6 and CA7 could be pseudogenes or expressed at very low levels or under specific conditions. CA1 (CAH1) was found in all organs except root, while CA2 was only expressed in stem and root and CA3 restricted to flowers and siliques (Fabre et al., 2007). CA3 has previously been found in mature pollen in proteomic analysis (Holmes-Davis et al., 2005; Noir et al., 2005). In addition, a cDNA for CA7 was isolated from an expanded Arabidopsis library (Yamada et al., 2003) and CA4 was identified in the thylakoid membranes in a mass-spectrometric proteomic approach (Friso et al., 2004; Sun et al., 2009). Sequence analysis shows that CA2 is lacking N-terminal targeting information, presumably encoding a cytoplasmic variant of the protein, while the remaining α-CAs have predicted SPs for co-translational insertion into the ER lumen.
In this chapter, results and conclusions from the paper and the manuscripts of the thesis are outlined and discussed in an attempt to summarize the different parts into one consecutive story. At the time of the initiation of this PhD project, a carbonic anhydrase, CAH1, had been found in the *Arabidopsis thaliana* chloroplast. What made CAH1 peculiar compared to other chloroplast proteins was the presence of N-linked glycans on the protein. N-glycosylation is only known to occur in the endomembrane system, and no route for proteins between the endomembrane system and the chloroplast was known at that time. In paper I this finding was reported, that a chloroplast stroma localized protein, instead of following the canonical route through the Toc-Tic system in the chloroplast envelope, was trafficking via the ER and Golgi to the chloroplast. Additionally, we could show that the protein arriving at the chloroplast was N-glycosylated. In an attempt to genetically examine components of the trafficking mechanism reported in paper I, manuscript II shows that proteins known for their involvement in canonical ER-to-Golgi vesicle trafficking also are involved in transport of CAH1. In addition, the manuscript describes an improved method for transient co-expression of multiple genes in plant cells. We also try to emphasize the potential of this optimized method and how it can be used in trafficking studies of CAH1 and/or other proteins. Manuscript III focuses on the importance of post-translational modifications of CAH1, and the effect of these modifications on fundamental processes such as folding, trafficking, and protein functionality and/or activity. The results from this study encourage us to present a hypothesis to why a protein trafficking pathway from the endomembrane system to the chloroplast exists in the plant cell. Finally, manuscript IV presents data obtained from Arabidopsis plants with disrupted CAH1 gene expression, which clearly indicate the relevance of the activity of this CA in the photosynthetic performance and the chloroplast function of a C3 plant.

### 2.1 Paper I

CAH1 was identified in a proteomic screening for chloroplast localized CAs in the plant model *Arabidopsis thaliana*. Chloroplast localization of the protein was suggested from subcellular fractionation of leaf material (Paper I, Figure 1d). Neither CAH1 gene transcript, nor the CAH1 protein itself, was detected
in roots, indicating that the protein is only present in above ground organs, presumably concentrated in leaf tissues. Electron microscopy of immunogold labelled cell sections confirmed that the CAH1 protein was localized in the stroma, the soluble interior of the chloroplast (Paper I, Figure 1c).

### 2.1.1 CAH1 is lacking N-terminal transit peptide for targeting to the chloroplast

Analysis of the primary sequence (amino acid sequence) can supply important information about a protein. For example, conserved domains of enzyme classes can be found, specific sites for post-translational modifications can be predicted and the secondary structure for different stretches of the polypeptide sequence can be deduced. Even the complete 3D structure of the protein can be modelled if structure of a related protein has been resolved. Additionally, comparison to the vast genome and protein sequence information now available from many organisms can give valuable clues about the function or localization of the protein.

One of the tools available for analysis is the algorithm known as TargetP (Emanuelsson et al., 2000), a neural network-based service for large-scale prediction of subcellular location. Targeting information is often present at the N-terminus of polypeptides, e.g. targeting to the secretory pathway, the mitochondria and, in plant cells, the chloroplasts. The overall success rate for analysis of protein targeting to these organelles is as high as 85%, in which proteins destined to the ER are most successfully predicted (95%) and chloroplast proteins being most difficult (69%). Analysis of the 284 amino acid long polypeptide of CAH1 strongly suggested that the protein was targeted to the ER, a prediction in total disagreement with the experimentally verified subcellular location of the protein. Additionally, TargetP proposed a cleavage site for the ER SP at amino acid 24 and 25 (ADA-QT, Paper I, Figure 1a). Although prediction of chloroplast located proteins is difficult, further analysis using Predotar (Small et al., 2004) confirmed suggested ER targeting.

Comparison to other CAs identified the region spanning from amino acid 35 to 262 as the carbonic anhydrase domain (PD000865) (Servant et al., 2002), and confirmed that the CAH1 polypeptide possessed an N-terminal extension not necessarily important for CA activity. While this analysis only was based on sequence analyses, it was important to determine whether the CAH1 precursor protein could be chloroplast imported, indicating false ER SP prediction. In vitro uptake studies were performed with isolated pea chloroplasts and dog pancreas microsomes to see if the CAH1 precursor was competent for import to any of the two cellular compartments. Expression of the protein in the presence of pea chloroplasts, often used in this type of analyses, showed that neither the CAH1 precursor, nor the polypeptide lacking the predicted ER SP, was imported (Paper I, Figure S1e). Instead, full length CAH1 was
efficiently translocated into dog pancreas microsomes (Paper I, Figure 2c). The polypeptide was additionally processed by a signal peptidase in the ER lumen, indicating that the CAH1 precursor protein indeed possessed a functional ER SP despite its chloroplast localization in planta. In addition, the protein accumulated in the microsomes with a similar weight as the native stromal protein (around 38 kDa) which is substantially larger than the expected mass of the polypeptide backbone alone (32.7 or 30.0 kDa, with or without SP, respectively). This difference in mass suggested that the mature CAH1 harboured some kind of post-translational modification.

At this point our experimental data were pointing in two different directions. Location of the protein appeared to be within the chloroplast while targeting information strongly indicated a protein heading for the endomembrane system, and not the chloroplast. The only way for us to put these observations together required a (novel) route for proteins from the secretory system to the chloroplast (Figure 2.1).

![Figure 2.1. New glycoprotein pathway to the chloroplast via the endomembrane system. CAH1 is translocated into the ER, where the protein acquires N-linked glycans. The N-glycans are processed as the protein travels from the ER to the Golgi. CAH1 is then transported, presumably via Golgi derived vesicles, to the chloroplast where the vesicles fuse with the outer envelope membrane. CAH1 finally crosses the envelope membrane by an unknown mechanism and reaches the stroma.](image)

As mentioned earlier, engulfment of a Gram-negative prokaryotic cyanobacterium with two membranes, by a eukaryotic cell surrounded by a plasma membrane, intuitively would have resulted in an endosymbiotic organelle with
three membranes (Figure 1.3). Biochemical analyses of chloroplast membranes have shown a chimeric outer envelope membrane, containing a mix of both prokaryotic and eukaryotic components and suggesting that the outer membrane originates from a fusion of two membranes (Kilian and Kroth, 2003). Presumably these two membranes correspond to the outer membrane of the cyanobacterium and the vacuolar membrane of the eukaryote. Since targeting of nuclear encoded chloroplast proteins requires an existing and functional import system, it is not straightforward to understand how the genomic rearrangement of cyanobacterial genes to the nucleus proceeded. If genes encoding cyanobacterial proteins were transferred to the nuclear genome before evolution of the Toc-Tic system, the resulting gene products would not have been imported. On the other hand, there would be no evolutionary pressure for a Toc-Tic system to arise if there were no proteins to import.

One proposed explanation to this enigma is that initial gene transfer resulted in early "chloroplast" proteins possessing SPs for the secretory system. If the outer envelope membrane contained eukaryotic components, either at a third membrane or at a chimeric secondary membrane, the receptors and factors for fusion of secretory vesicles could exist in such membrane. Secretion of endosymbiotic proteins could then additionally be targeted to the early chloroplast, while a functional Toc-Tic system developed (Kilian and Kroth, 2003).

Organisms where secondary or tertiary endosymbiosis has occurred, i.e. where a non-photosynthetic eukaryotic host cell engulfed a photosynthetic eukaryote (a green or red alga), contain secondary plastids surrounded by more than two, usually three or four, membranes (Kilian and Kroth, 2003; Raven and Allen, 2003; Sanchez-Puerta and Delwiche, 2008). At least three secondary endosymbiotic events are recognized today (Sanchez-Puerta and Delwiche, 2008), two involving a green algae (leading to plastids of euglenoids and chlorarachniophytes) and a third involving a red alga giving rise to plastids of different eukaryotic lineages (haptophytes, heterokonts, cryptophytes, dinoflagellates and apicomplexa) (Keeling, 2009). Some of these plastids have retained the photosynthetic capacity, while others have not. An example of an organism with secondary plastid no longer capable of photosynthesis is the apicomplexan parasite *Plasmodium falciparum* (Kilian and Kroth, 2003) (Figure 1.3). *P. falciparum* possesses cytoplasmic organelles with three membranes (called apicoplasts). Intriguingly, import of proteins into these apicoplasts is a two-step process where a SP is mediating import into the endomembrane system, where it is cleaved off, revealing a TP that diverts the protein from the secretory pathway to the apicoplast (Foth et al., 2003). In a similar way, proteins destined to secondary plastids in diatoms contain bipartite pre-sequences with a SP followed by a TP. The SP of these proteins has also been demonstrated to be functionally equivalent to precursor sequences of normal ER-targeted proteins (Kilian and Kroth, 2005).

The presence of a protein pathway to the chloroplast through the secretory system in a higher plant would conclusively demonstrate that chloroplast
protein import is not exclusively dependent on a functional Toc-Tic system. In addition, such a finding could suggest that sorting of proteins to the evolving chloroplast initially might have occurred via the secretory system (Reyes-Prieto et al., 2007), in a similar way as for secondary plastids of diatoms and *P. falciparum*.

### 2.1.2 CAH1 contains an N-terminal signal peptide for the ER, but is localized to the chloroplast

To verify chloroplast targeting of the protein by a method independent of the use of CAH1 antibodies, a translational fusion of CAH1 and GFP (green fluorescent protein) was constructed. GFP was fused C-terminally to CAH1 to avoid interfering with the ER precursor sequence of CAH1. As expected from immunogold and subcellular fractionation experiments, the protein localized to the chloroplast as seen by confocal microscopy (Paper I, Figure 1e). In contrast, addition of an ER retention signal (KDEL) to the C-terminus of the CAH1-GFP construct resulted in fusion protein being localized to the ER (Paper I, Figure 2a and b). The KDEL tail was added to ensure that fusion protein translocated into the ER lumen would be retained and not further trafficked. No GFP signal was detected in the chloroplast, excluding the possibility that the protein was simultaneously sent to ER and chloroplast by dual targeting, a phenomenon that had been reported in previous publications (Levitan et al., 2005). This result concluded that CAH1 indeed contained active and functional precursor sequences for the ER an ER only.

CAH1 is presumed to be a low-abundance protein in Arabidopsis, and efforts to deduce the exact signal peptidase cleavage site by N-terminal sequencing of the native protein failed. Instead, another approach was tested in which the gene construct for the ER-retained GFP fusion protein was stably transformed and expressed in Tobacco BY2 cell suspension culture. Accumulation of the protein in these cells proved to be high, and ER-retained protein harvested from the total extract and purified by a single step of anion exchange chromatography could be N-terminally sequenced. Since processing in the ER is highly conserved, the cleavage site in the native protein could be assumed to be identical to the ER retained polypeptide (ADAQ-T), notably only one residue from the site predicted by SignalP (Bendtsen et al., 2004).

### 2.1.3 Chloroplast localized CAH1 is N-glycosylated in the ER

A common post-translational modification taking place in the ER of eukaryotic cells is N-glycosylation. N-glycosylation results in sugar complexes anchored to specific sites of polypeptides (Asn-X-Ser/Thr, where X can be any
amino acid but Pro). CAH1 has five such sites, of which four are predicted to be decorated with N-linked glycans (www.cbs.dtu.dk/services/NetNGlyc). Each N-glycan has a mass of about 1.5-2 kDa, depending on the number and type of sugar molecules present, resulting in a theoretical increase of CAH1 weight by 6-10 kDa, well in accordance to the 8 kDa difference seen between SP processed protein and mature stroma CAH1 (30.0 and 38 kDa, respectively) (Paper I, Figure 2c, lane 4 and Figure 3b). Inhibition of N-glycosylation during uptake studies into dog pancreas microsomes confirmed that the protein was glycosylated in vitro and that the glycosylated polypeptide migrated with a similar weight as the native protein, strongly suggesting that the chloroplast located protein was glycosylated.

Targeting of CAH1 to the ER was shown both in vitro and in vivo using uptake studies into microsomes and confocal microscopy of KDEL tagged protein, respectively, but this did not say anything about further trafficking of the protein from the ER to the chloroplast. Translocation of the protein out of the ER, followed by Toc-Tic mediated translocation, seemed unlikely since this would require unfolding of the protein for passage through the envelope translocon. Also, the presence of bulky N-glycans would certainly affect transport through Toc-Tic. Other scenarios could mimic the targeting of proteins to the secondary plastids. In order to avoid passage through vacuoles, the two most likely alternatives would be direct targeting from the ER to the chloroplast, or via the Golgi apparatus.

The most predominant forms of lipids in the chloroplasts are galactolipids. While assembly of galactolipids takes place in the chloroplast envelope, the galactolipid precursors (diacylglycerol moieties) of many plants species (including Arabidopsis) originate from two different compartments, the plastid (prokaryotic pathway) or the ER (eukaryotic pathway). In the prokaryotic pathway diacylglycerol is assembled directly from fatty acids synthesised in chloroplasts and incorporated into chloroplast galactolipids. In the eukaryotic pathway the fatty acids are transported to the ER where diacylglycerol is assembled and later returned to the chloroplast envelope for galactolipid synthesis (Benning et al., 2006; Kelly and Dormann, 2004; Xu et al., 2008). Obviously, such events require some kind of interaction between the ER and the chloroplast. Not only have contact sites between ER and plastids been reported, so-called plastid associated membranes (PLAMs) (Hanson and Köhler, 2001; Kunst and Samuels, 2003), but protein-protein interactions have also been suggested (Andersson et al., 2007) and recently a protein responsible for mediating lipid transfer between the ER and the outer envelope membrane was suggested (Xu et al., 2008). Considering reported interactions between plastids and the ER, a direct transfer from the ER to the chloroplast emerged as an attractive explanation for transport of CAH1.
2.1.4 Attachment of complex type glycans to CAH1 suggest a route to the chloroplast via the Golgi apparatus

One way to study targeting of glycoproteins is to analyse the type of N-linked glycan attached to the protein. Maturation of N-glycans along the secretory pathway has previously been described and the subcellular action of several glucosidases and glucosyl transferases determined (Figure 1.7). For example, glucosidases I and II remove the glucose residues of the oligosaccharide precursor Glc3Man9GlcNAc2 in the ER (Lerouge et al., 1998), resulting in a structure known as high mannose-type N-glycans. These N-glycans can be further modified in the Golgi, resulting in complex-type N-glycans. Plant complex-type glycans are characterized by the presence of α(1,3)-fucose and/or a β(1,2)-xylose residues, respectively, linked to the proximal N-acetyl-glucosamine (GlcNAc) residue of the chitobiose core and the β-mannose residue of the core, and by the presence of β(1,2)-GlcNAc residues linked to the α-mannose units (Lerouge et al., 1998; Rayon et al., 1998).

By analyzing the type of N-linked glycans attached to the CAH1 protein we could deduce whether the protein was targeted directly to the chloroplast from the ER, or if CAH1 was trafficking via the Golgi. Immunoprecipitation (IP) of CAH1-GFP fusion protein and subsequent Western blot analysis using antibodies against α(1,3)-fucose (Paper I, Figure 3a) clearly showed presence of complex type N-glycans anchored to the chloroplast localized polypeptide. This finding that complex type glycans are attached to CAH1-GFP proved that the protein was trafficking through the Golgi apparatus on the way to the chloroplast. The presence of α(1,3)-fucose residues was not detected in the ER-retained CAH1-GFP-KDEL, indicating that the KDEL-tagged variant was not reaching the Golgi and that the antibody binding was specific.

To provide further evidence that the native protein was following the same route as the GFP-tagged polypeptide, proteins from Arabidopsis stroma was separated by 2D-gel electrophoresis. Western blot analysis using antibodies against CAH1, α(1,3)-fucose or β(1,2)-xylose, together detected a series of spots that appeared to originate from the same polypeptide, indicating that the native protein also harboured complex type N-glycans (Paper I, Figure S2). Attachment of complex type N-glycans, containing both α(1,3)-fucose and β(1,2)-xylose residues, to native CAH1 was later confirmed by enzymatic treatments. Endo H, which only removes high mannose type glycans, did not change the migration pattern of CAH1. In addition, CAH1 protein isolated from wild type (wt) Arabidopsis was resistant to peptide-N-glycosidase F (PNGase F), an enzyme that cannot cleave N-glycans containing α(1,3)-fucose residues. However, CAH1 isolated from the mur1 mutant was sensitive to PNGase (Paper I, Figure 3c). The mur1 mutant is unable to synthesize fucose (Bonin et al., 1997) therefore its endogenous N-glycoproteins do contain complex type N-glycans without α(1,3)-fucose residues which will be sensitive to PNGase F.
2.1.5 Trafficking of CAH1 is blocked by Brefeldin A

The presence of complex-type N-glycans anchored to the CAH1 polypeptide indicated that this protein is trafficking from the ER via the Golgi to the chloroplast. To test whether the protein could be blocked using established inhibitors of the trafficking pathways through the endomembrane system, protoplasts expressing chloroplast targeted CAH1-GFP were treated with brefeldin A (BFA). BFA is an antibiotic produced by fungal organisms that interferes with the function of the GTPase Arf1 in COPI coat recruitment, eventually leading to an integration of the ER and Golgi apparatus (Baluska et al., 2002; Batoko et al., 2000; Brandizzi et al., 2002; daSilva et al., 2004; Xu and Scheres, 2005).

Addition of BFA caused an accumulation of the fluorescently tagged protein in large aggregates, similar to the effect of the BFA treatment on trafficking of well-known secretory proteins (Lee et al., 2002). When the BFA was washed away, normal ER and Golgi structures could be reformed whereby the protein was released and finally detected in the chloroplast (Paper I, Figure 4a-e). Addition of the protein biosynthesis inhibitor cycloheximide (CHX) when BFA was washed away ensured that the same protein seen blocked upon BFA treatment was later on detected in the chloroplasts.

2.2 Manuscript II

To continue our study of the mechanisms for transport of CAH1 between the ER and the Golgi, we wished to explore the requirement for specific GTPases involved in vesicle trafficking. Although BFA arrested CAH1-GFP in aggregate-like structures, BFA is known to have different effects in different species, and even in different tissues within a species (Robinson et al., 2008). In addition, BFA at high concentrations is likely to induce secondary effects, and the effect of BFA on trafficking of CAH1 should be seen only as a first indication of COPII mediated vesicle transport between the ER and Golgi.

Involvement of three individual GTPases was tested: Sar1, RabD2a, and Arf1, all of them playing important roles in vesicle formation or docking at the ER/Golgi interface. Single site mutagenesis of these GTPases has resulted in arrested and non-functional enzymes. RabD2a is a small GTPase involved in targeting and fusion of ER-derived COPII vesicles at the Golgi surface. Dominant negative variants of the protein, where an N121I (Asn121 to Ile) substitution was introduced in the GTP binding motif, was created. This variant was shown to have inhibited trafficking of secreted and Golgi targeted proteins out from the ER (Batoko et al., 2000; Jurgens, 2004; Pinheiro et al., 2009). On the contrary, Sar1 and Arf1 are directly involved in the formation of COPII and COPI vesicles, respectively (Jurgens, 2004; Takeuchi et al., 2002).
Two mutant isoforms of Arf1 were tested: Arf1(Q71L) and Arf1(T31N), both shown to affect ER-to-Golgi trafficking and to relocate Golgi markers to the ER (Takeuchi et al., 2002). The Q71L (Gln71 to Leu) mutant shows reduced GTPase activity, therefore acting as a constitutively activated mutant interfering with sorting of membrane proteins into Golgi-derived COPI vesicles. The T31N (Thr31 to Asn) mutant instead has low affinity for GTP, acting as a dominant-negative mutant that blocks formation of COPI vesicles (Xu and Scheres, 2005). The mutant version H74L (His74 to Leu) of the Sar1 small GTPase protein results in a dominant mutated variant which is insensitive to its GTPase-activating factor and thus it is fixed as a GTP-bound form. This mutant variant changes the localization of cis-Golgi marker to the ER as well as blocks trafficking of a vacuolar protein out of the ER (Takeuchi et al., 2000). These mutant proteins are believed to be dominant by titrating out factors needed for normal GTPase activity, and therefore out compete the function of the native proteins.

To study the effect of these dominant mutant GTPases on trafficking of CAH1, we wished to use an experimental system in which expression of the mutant proteins could be controlled by an inducible system. Arf1 expression, induced by the onset of a heat shock promoter (Xu and Scheres, 2005), in four-week-old plants grown on soil was tested. Unfortunately induction in leaves was very low, which in combination with the preparation time needed for good chloroplast isolations, convinced us that a different experimental system had to be found. In addition, disrupted protein biosynthesis by CHX indicated that turnover of the CAH1 protein was low. If these GTPases affected trafficking, low turnover could result in a concealed inhibitory effect, where protein already located at the chloroplast would remain at nearly wt levels and potentially mask blocked chloroplast targeting by the mutant GTPases.

In conclusion, a system that could be transiently induced was desired since constant expression of these mutants likely would be lethal for the plant. At the same time, the experimental setup had to ensure that only CAH1 expressed following induction of the mutant GTPase was analyzed. Last but not the least, the CAH1 that was being studied must originate from the same cell as that where the mutant GTPase was expressed. The last requirement is a problem when using transient expression techniques, such as transfection of protoplasts, where only a fraction of the cells are transfected while the majority are behaving as wt cells. One way to solve this was to use a fluorescently tagged mutant protein, and exclusively analyze cells exhibiting fluorescence. Unfortunately, previous studies on some GTPase proteins have reported that the proteins might become inactive when a tag is present (Samalova et al., 2006).

To summarize, we were looking for an experimental system fulfilling these criteria:

1. Enabling transient expression of the mutant protein (GTPase) in the absence of any kind of tagging system.
2. Expression of the mutant protein must take place in the same cell as the CAH1 protein that is being analyzed.

3. The CAH1 protein that is being analyzed must have been expressed at the same time or following expression of the mutant GTPase.

2.2.1 Development of an optimized 2A peptide co-expression system

To fulfil these three criteria, we decided to optimize the use of a short peptide called 2A peptide. The 2A peptide is a 16-20 amino acid long peptide that is used by some RNA viruses and enables the synthesis of several gene products (proteins) from a single transcript without requiring extra-ribosomal factors (Donnelly et al., 2001; Halpin et al., 1999; Szymczak et al., 2004). During translation of the transcript, the 2A peptide causes a premature release of the polypeptide N-terminus of the 2A peptide without stopping translation of the transcript. The process is hereafter termed 'cleavage', although no protease is involved. In short, several individual proteins can be produced from one transcript consisting of several genes joined by the sequence encoding the 2A peptide. The 2A peptide system in combination with an epitope tagged version of CAH1 would fulfil all three criteria. Protoplast transfection with this construct would transiently express the studied GTPase, and expression of the GTPase would only take place in the same cell and at the same time as epitope tagged CAH1 (Figure 2.2).

The 2A peptide had previously been tested in plant cells, but efficient cleavage of the 2A polyprotein was questioned. A recent publication indicated that less than 50% of the polyprotein was actually cleaved (Samalova et al., 2006). As trafficking of CAH1 and function of the GTPase required individual polypeptides, 2A peptide cleavage efficiency had to be optimized before the effect of the mutant GTPases could be assessed. Initially, two factors potentially influencing cleavage efficiency were tested.

2. Primary sequence N-terminal of the 2A peptide.

A fluorescent marker protein (CFP), transcriptionally fused to RabD2a via the 2A peptide (CFP-2A-RabD2a), constituted the starting point used in the development of an efficient 2A expression system. CFP facilitated visualization in vivo and cleavage efficiency could be quantified by Western blotting using commercially available antibodies. Antibodies detecting RabD2a GTPase, kindly provided by Dr. Ian Moore (Oxford University, UK) were used to verify that separation of the polypeptides was due to 2A peptide cleavage, instead of protease degradation of non-cleaved polyprotein, and that the separated GTPase was stable after separation from the 2A peptide polyprotein.
2A peptide constructs translated by ER bound ribosomes results in increased cleavage efficiency. Expression by cytosolic ribosomes resulted in low cleavage efficiency (about 20%) (Manuscript II, Figure 2h), in accordance with previous expression assays of 2A peptide constructs in plant cells (Samalova et al., 2006). Directing translation to the ER membrane using the ER SP of CAH1 (SPCFP-2A-RabD2a) or the Golgi marker NAG (NAGCFP-2A-RabD2a, Grebe2003) increased cleavage efficiency considerably (to 55-70%). In addition, insertion of the β-glucuronidase (GUS) protein between CFP and the 2A peptide (CFPGUS-2A-RabD2a) to alter the N-terminal sequence of the 2A peptide surprisingly increased the cleavage efficiency even more (to over 90%), indicating that several (unexpected) factors influence cleavage efficiency of 2A polyprotein (data not shown).

To demonstrate that cleavage efficiency was not affected by differences in stability of the 2A polyprotein or separated polypeptides, transfected protoplasts were incubated with CHX and the ratio between non-cleaved and cleaved 2A polyprotein at different time points was estimated. Rapid degradation of non-cleaved 2A polyprotein, due to instability or aggregate formation, would result in low levels of non-cleaved polyprotein and produce an overestimation of the 2A peptide cleavage efficiency. Importantly, the ratio of non-cleaved polyprotein to cleaved polypeptide did not alter upon CHX treatment, indicating that the estimated cleavage efficiencies were likely to be correct (data not shown).
2.2.2 Released GTPase proteins were functional

Western blot analysis of RabD2a cleaved from 2A polyprotein migrated with a similar weight and at comparable amounts as non-tagged and over-expressed RabD2a (Manuscript II, Figure 2S), confirming that separation of polypeptides from the 2A polyprotein was specific and not due to proteolytic processing of the polyprotein. It also appeared that the released RabD2a showed similar turnover to the native protein, although an extra proline residue was present in the N-terminus of 2A peptide-derived RabD2a (as a result of cleavage of the 2A peptide).

Up to this point, successful development of the 2A peptide system resulted in increased cleavage efficiency and separated polypeptides showing similar stability as their native variants. What remained was to show that the 2A peptide-derived protein was functional and that the residual Pro residue did not interfere with the inhibitory effect of the mutants. Protein location is crucial for its function and the GTPases studied required cytosolic localization. Since translation by ER membrane-bound ribosomes appeared more efficient than by free ribosomes, release of the 2A peptide-derived GTPase had to take place before the GTPase was translocated across the ER membrane (i.e. release had to take place at the cytosolic side of the membrane).

Confocal microscopy of the subcellular location of the fluorescent CFP protein showed the expected location for all four tested constructs (CFP-2A/CFP-GUS-2A/SPCFP-2A/NAGCFP-2A). CFP-2A and CFPGUS-2A were cytosolic, while SPCFP-2A and NAGCFP-2A showed membrane-bound localization where SPCFP-2A was mainly seen in the ER and NAGCFP-2A in the Golgi (Manuscript II, Figure 2a-f). Functionality of the 2A peptide-derived GTPases was proven from the effect on the subcellular localization of the NAGCFP Golgi marker. NAGCFP-2A, co-expressed with wt isoforms of RabD2a, Sar1 or Arf1, showed expected and distinct Golgi localization. On the contrary, co-expression with the mutant variants of the three GTPases resulted in reticulated fluorescence, clearly indicating that the 2A peptide-derived mutant GTPases inhibited targeting of the Golgi marker and that vesicle trafficking between the ER and the Golgi was disturbed (Manuscript II, Figure 3 and 4). Functionality not only indicated that the cleavage was efficient, but also that the separated polypeptides were correctly targeted to their individual destination.

All results pointed to a successful development of the 2A peptide co-expression system, and showed that it could be used to simultaneously express several proteins from one transcript.
2.2.3 Endo H resistance assays can reveal information about trafficking

Trafficking of N-glycoproteins from the ER to the Golgi results in maturation and processing of the N-linked glycan, converting the glycoprotein from Endo H sensitive to Endo H resistant glycoform. This has been proven to be a fast, easy and efficient assay to study trafficking of N-glycoproteins from the ER (Downing et al., 2006; Hong et al., 2008). As long as the protein is retained in the ER, the protein harbours N-linked glycans sensitive to Endo H, while ER-export and trafficking to the Golgi exposes the protein to mannosidase II, inferring resistance to Endo H (Figure 1.7). Separation of the polypeptide by SDS-PAGE after Endo H treatment can give important information about the attached N-glycan, and also about the subcellular localization and trafficking of the protein.

We decided to combine the Endo H assay with our optimized 2A peptide co-expression system to answer whether these GTPases are involved in CAH1 trafficking. The fluorescent marker protein was replaced by an HA-tagged variant of wt CAH1 (HA-CAH1, Manuscript II, Figure 1), enabling simultaneous expression of HA-CAH1 with wt and mutant variants of RabD2a, Sar1 or Arf1. Stable expression of HA-CAH1 in suspension cultured cells confirmed that the HA-tagged protein was targeted to the chloroplast and that the protein was N-glycosylated (Manuscript II, Table I and Figure 6). Since cleavage efficiency was higher when translation was performed by ER membrane bound ribosomes, and HA-CAH1 pre-protein possesses a SP for the ER, the HA-CAH1 protein was inserted N-terminal of the 2A peptide (generating HA-CAH1-2A-GTPase) (Manuscript II, Figure 1). Cleavage efficiency was high (about 85%), confirming that translation by ER bound ribosomes in our system results in high cleavage efficiency.

In an attempt to quantify the effect of the different variants of the mutant GTPases on the trafficking of HA-CAH1 protein, the ratio of Endo H resistant to Endo H sensitive HA-CAH1-2A was calculated. A clear mutant phenotype was revealed from co-expression of mutant Rab2a and Sar1 (Manuscript II, Figure 5), as anticipated from their direct involvement in ER-to-Golgi trafficking. Arf1, involved in retrograde trafficking from the Golgi to the ER, resulted in a less strong phenotype. While there seemed to be an effect of the GDP-locked variant, no significant effect could be deduced from GTP-locked Arf1. Interestingly, a similar effect was presented in a recent publication where trafficking of rice α-amylase isoform I-1 (AmyI-1) to the plastid in onion epidermal cells was evaluated in the presence of mutant Sar1 and Arf1 variants (Kitajima et al., 2009). They could show that mutant Sar1 strongly inhibited arrival of the AmyI-1 protein to the plastid, while Arf1 had a smaller inhibitory effect. Also in their system, the GDP-locked Arf1 variant gave clearer mutant phenotype than the GTP-locked isoform (Kitajima et al., 2009).
The paper by the Japanese group (Kitajima et al., 2009) reported about the trafficking mechanism of an N-glycosylated protein to the plastid in monocot systems (rice, *Oryza sativa*, and onion, *Allium cepa*). Their results were in good agreement with our Endo H assay, where involvement of RabD2a and Sar1 in trafficking of CAH1 was obvious. In a previous study, they published data regarding another rice plastid enzyme, a nucleotide pyrophosphate/phosphodiesterase 1 (NPP1). This protein was also targeted to the plastid through the endomembrane system. Moreover, the mature form of NPP1 was N-glycosylated. Interestingly, and to us unexpectedly, the plastid localized proteins NPP1 and Amy1-I seemed to harbour high mannose-type N-glycans, as revealed by binding of two glycoproteins to the lectin concanavalin A (ConA) and their susceptibility to Endo H (Asatsuma et al., 2005; Nanjo et al., 2006), indicating either a direct transport of these two proteins from the ER to the plastid, or that the N-linked glycans were not further processed in the Golgi. The occurrence of a rice-specific glycosylation pattern can be ruled out since the monocot N-glycosylation machinery has been shown to be identical to the one found in Brassicacea family of dicot plants (Léonard et al., 2004). Closer evaluation of the NPP1 plastid localization indicated that, in addition to uniform internal plastid labelling, numerous distinct dots of NPP1-GFP, seemingly located at the plastid envelope, were visible (Nanjo et al. 2006, Figure 4). In our eyes, these dots strongly resembled ER-chloroplast contact sites, PLAMs, previously reported (Andersson et al., 2007). Based on this, one could speculate about a direct transfer of NPP1 from the endomembrane system to the plastid by a similar mechanism as the transport of lipid precursors from the ER to the plastid (Benning et al., 2006; Kelly and Dormann, 2004; Xu et al., 2008).

In an attempt to study targeting of NPP1 in our protoplast system, rice NPP1 was fused to GFP and transfected into Arabidopsis protoplasts. The fusion protein was expressed at significant levels. In addition, it was sensitive to Endo H, indicating that the polypeptide acquired high mannose-type N-glycans as when expressed in the native rice system. Despite these facts, we could not verify plastid targeting in Arabidopsis protoplasts. Instead NPP1-GFP appeared to accumulate in the ER (data not shown), indicating that rice NPP1 is not targeted to chloroplasts in a dicotyledon plant such as *Arabidopsis thaliana*, or that targeting is rice or monocot specific.

### 2.2.4 Potential areas of application for the 2A peptide co-expression system

One drawback with transient expression techniques, such as protoplast transfection, is that when a visible phenotype is lacking, analysis of your sample is based on a mixture of transfected and non-transfected cells (mutant and wt cells). If either the effect one aims to study, or the transfection efficiency is very low (or a combination of both), the resulting phenotype might be
Figure 2.3. Obtaining a homogenous mutant protoplast population using FACS. Transfection of protoplasts results in a heterogeneous population where only a fraction of the protoplasts are actually transfected. Analysis of the protoplasts after transfection (Step 2) shows that the transfected population (T), in contrast to the non-transfected cells (NT), is expressing the transgene. While the protein extract in step 2 originates from a mix of transfected and non-transfected cells, further sorting of the transfected population based on GFP fluorescence results in a homogenous mutant population (Step 3). Western blotting using anti-BiP (a chaperone in the ER) or anti-GFP anti shows that only the transfected population is expressing GFP (Step 2), and that GFP fluorescing protoplasts can be efficiently sorted (Step 3).
too small to be detected among the background generated by non-transfected cells. In such cases it would be desirable to sort out mutant cells in order to achieve a homogenous mutant population. To test whether the 2A system offered such possibilities, protoplasts were transfected with the Golgi labelled NAGCFP-2A-RabD2a construct and analyzed by flow cytometry, using a fluorescence-activated cell sorter (FACS). A subpopulation corresponding to about 10% of the intact protoplasts was defined as positively transfected based on their level of CFP fluorescence. Protoplasts belonging to both the transfected and non-transfected group were sorted and examined by fluorescence microscopy. A clear CFP fluorescence could be detected in the positively sorted protoplasts. In addition, Western blot analysis of both populations revealed a strong CFP signal in the extract from the positively sorted population, while no signal could be detected in extract from the negatively sorted protoplasts (Figure 2.3).

In conclusion, the versatility of the 2A peptide co-expression system offers several possibilities for over-expression of proteins where direct tagging is not possible. Co-expression with fluorescent markers can in such cases enable easy visualization of cells in which the protein is being expressed. Also, the effect derived from the over-expression of the protein can be studied in real time using different microscopic techniques, such as confocal microscopy. Furthermore, sorting of transfected cells using FACS, can allow obtaining of a homogenous mutant population expressing your favourite protein in less than 24 h.

2.3 Manuscript III

While paper I and manuscript II focused on localization and trafficking of CAH1 to the chloroplast, one obvious question remained to be considered - why CAH1 is targeted to the chloroplast via the secretory pathway instead of the Toc-Tic translocation system from the cytosol as for most other chloroplast proteins. We were also lacking important biochemical information about the protein, in particular that related to post-translational modifications. In paper I we reported that the protein is N-glycosylated. N-glycosylation is exclusively initiated in the ER and constitutes a logical starting point for further analysis of the protein. Over-expression of epitope-tagged versions of CAH1 in different systems -BY2 Tobacco and Arabidopsis cell suspension cultures, as well as in Arabidopsis and Tobacco plants- always resulted in its accumulation at rather low amounts, hampering the purification of CAH1 protein to substantial quantities. So, instead of a mass spectrometric approach, we decided to do mutagenesis of the protein. A collection of amino acids, mainly representing targets for N-glycosylation, were point mutated and the resulting mutated protein versions expressed in protoplasts.
2.3.1 CAH1 harbours four or five N-glycans

First we wanted to elucidate the number and positions of N-glycans anchored to the protein. As seen in paper I, theoretical and native molecular masses of CAH1 differed by approximately 8 kDa, indicating that around 4-5 N-glycans could be decorating the CAH1 protein (supposing about 1.5-2 kDa/N-glycan). All five sites were point mutated, both individually and in combination, resulting in single, double, triple, and quadruple mutants in addition to a non-glycosylated variant (Manuscript III, Table 1). All mutant proteins possessed an HA-epitope tag to enable the use of the highly specific anti-HA antibody and to avoid detection of native protein (as background).

It appeared as if the vast majority of the HA-CAH1 molecules were harbouring four or five N-glycans, and that partial occupation of glycosylation site four (Asn194) resulted in these two isoforms (Manuscript III, Figure 2 b and c). This result was in agreement with the in vitro import study into microsomes in paper I, where two dominating bands were seen (Paper I, Figure 2c). In addition, we detected glycoforms containing either high mannose- and complex type N-glycans of both 4- and 5-glycan HA-CAH1, explaining the four-band-pattern seen when the over-expressed HA-CAH1 from protoplasts was separated in SDS-PAGE (Manuscript III, Figure 3 a-c). Interestingly, two isoforms of the native protein, exhibiting slightly different molecular masses, were also frequently detected in leaf extracts, presumably indicating that CAH1 can also accumulate as two glycoforms in planta. None of these two glycoforms could be deglycosylated by Endo H, indicating that both harbour complex type N-glycans (data not shown). Moreover, these results suggest that partial occupancy of N-glycosylation site four is an intrinsic characteristic of the CAH1 protein, and not a secondary effect of the over-expression of the tagged isoform. While the heavier isoform was prevailing in some protein preparations, the opposite could been seen in other preparations. The reason for this behaviour is unknown to us, but could depend on some kind of regulation of the protein. N-glycosylation site four (Asn194) is adjacent to a cysteine residue (Cys191). Corresponding cysteine residue in human α-type CA IV has been reported to form part of an intramolecular disulphide bridge (Waheed et al., 1996). If a similar cysteine bond would appear in CAH1, the presence of N-linked glycan at Asn194 could interfere with bridge formation and result in synthesis of an unstable or non-functional CAH1, alternatively producing an enzyme with altered activity.

The presence of both high mannose and complex type N-glycans seemed to originate from over-expression of the HA-CAH1 protein, perhaps due to saturation of the CAH1 targeting machinery. It was recently shown that protein trafficking from the Golgi apparatus to the plastid in onion epidermal cells was tightly regulated (Kitajima et al., 2009). Inhibition of protein biosynthesis by CHX reduced the fraction of high mannose-isofoms and suggested that glycoproteins harbouring high mannose-type N-glycans originated from newly synthesized HA-CAH1 molecules waiting to be trafficked from the ER to the
2.3.2 A disulphide bridge is important for folding and ER-export of CAH1

Disulphide bridges between cysteine residues are important for protein stability as well as for regulation of enzyme activity. The ER is a highly oxidizing environment for proteins, and several ER chaperones assist in formation of such bonds (Hebert and Molinari, 2007; Jensen, 2006). Sequence analysis of CAH1 revealed four cysteine residues in the SP processed polypeptide, of which the first and the third (Cys27 and Cys191) were positioned close to each other in the 3D-structure modelled protein, indicating that a similar disulphide bridge could exist in CAH1 as in human CAIV (Waheed et al., 1996) (Manuscript III, Supplemental Figure 7). Separation of the native protein under reducing and non-reducing conditions demonstrated altered migration pattern and that CAH1 was sensitive to reducing agents, such as β-mercaptoethanol (β-ME) (Manuscript III, Figure 5). Mutating all four cysteine residues individually generated two isoforms apparently resistant to 2-ME. As expected, they corresponded to Cys27 and Cys191 and suggested a disulphide bridge between these residues (Manuscript III, Figure 6 a). Formation of these bridges in the ER is a crucial part of the folding process. If a protein cannot fold correctly it will be retained in the ER, bound to a chaperone (BiP) and then translocated over the ER membrane for degradation by the proteasome in the cytosol (Hebert and Molinari, 2007; Pimpl et al., 2006). Retention of a protein in the ER accordingly also prevents processing of its N-glycans in the Golgi, and the protein accumulates as Endo H sensitive glycoprotein. Therefore, an Endo H assay, similar to the one used in manuscript II to study effect on trafficking upon co-expression of mutant GTPases, could be employed in monitoring misfolding of the cysteine residue mutants. Indeed, mutating Cys27 and Cys191 resulted in a protein with significantly higher proportion of Endo H sensitive glycoforms compared to the wt protein, or isoforms where any of the other two cysteine residues (Cys68 and Cys229) were mutated (Manuscript III, Figure 6 d and e). The requirement of a disulphide bond between Cys27 and Cys191 for folding of ER-export competent protein was finally confirmed by the demonstrated association of the ER-chaperone BiP with proteins mutated at any of these sites (Manuscript III, Figure 6 b and c).
2.3.3 Non-glycosylated CAH1 forms aggregates

To our knowledge, chloroplasts lack N-glycosylation machinery. A functional requirement of N-linked glycans at CAH1 could explain why the protein is trafficked through the endomembrane system to the chloroplast. Similar to the requirement of disulphide bridges for correct structure of certain proteins, N-glycosylation has also been shown to be required for folding (Mitra et al., 2006; Parodi, 2000). Expression of the non-glycosylated mutant and separation of the resulting polypeptide by SDS-PAGE under non-reducing conditions resulted in the presence of an additional band around 65 kDa. This additional band presumably corresponded to a dimer of the non-glycosylated protein due to the formation of non-native inter-molecular disulphide bridges (Manuscript III, Figure 4a). In addition, several high molecular mass forms were detected by the HA-antibodies, indicating that aggregates of the protein were formed when no glycans could be attached to the HA-CAH1 polypeptide. Similar to the disulphide bridge mutants, misfolding of the non-glycosylated protein was verified by its increased association with BiP (Manuscript III, Figure 4b and c).

Lack of N-glycosylation or a correct disulphide bridge in CAH1 results in decreased CA activity. Activity of CAs can be estimated by measuring the rate by which carbon dioxide in CO\(_2\) (g) saturated water is converted to bicarbonate in a solution containing the CA enzyme. This inter-conversion lowers the pH of the solution and can be recorded using a fast and sensitive pH electrode. Measurements are often performed with purified proteins expressed in bacteria, but since bacteria (as chloroplasts) lack the N-glycosylation machinery, such expression would generate non-glycosylated CAH1. To overcome this, activity was instead measured on protein extracts from transfected protoplasts.

Expression of a cysteine mutant protein unable to form an intra-molecular disulphide bridge, and non-glycosylated HA-CAH1, resulted in mutant protein extracts with almost identical CA activity, significantly lower than the activity measured in the wt extracts. Considering the rather low transfection efficiency (about 5-10% as calculated from GFP transfected cells) and the dilution effect caused by having mostly non-transfected protoplast exhibiting their own CA activity, this experiment clearly indicated the importance of N-glycosylation and disulphide bridge formation for the assembly of a functional CAH1 (Manuscript III, Figure 7).

2.3.4 Evolutionary aspects of CAH1

The biochemical characterization of CAH1 showed that N-linked glycans, as well as disulphide bridge formation, were required for correct folding of the CAH1 protein. When not correctly folded, the protein appeared to be unable
to be exported from the ER and presumably was translocated out of the ER and degraded. We propose that these features of CAH1 explain why the protein is trafficked via the endomembrane system to the chloroplast.

The question that then arises is whether CAH1 is a protein with eukaryotic origin that found its way to the chloroplast where it acquired a new function within the endosymbiont, or whether CAH1 is a protein with cyanobacterial origin whose gene was transferred to the eukaryotic genome and later was transported back to the evolving chloroplast using the proposed, early route, for proteins (Bhattacharya et al., 2007; Reyes-Prieto et al., 2007). Blasting the CAH1 protein sequence against the NCBI database (www.ncbi.nlm.nih.gov) and excluding Viridiplantae from the search algorithm, proteobacterium proteins rank at the top. Proteobacterium is considered to be the ancestor of the mitochondrion (Cavalier-Smith, 2009), suggesting that CAH1 might have an origin from outside the cyanobacteria.

A proteomic study of Arabidopsis chloroplasts (Kleffmann et al., 2004), published in 2004 (before the pathway of CAH1 was discovered), identified 690 proteins. Unexpectedly for the authors, many of the identified proteins did not possess targeting information for the canonical Toc-Tic system, as shown using the TargetP prediction tool on the identified polypeptides (Emanuelsson et al., 2000). Subsequent to removal of envelope proteins, which often do not have canonical chloroplast TP (as discussed earlier), and other obvious contaminants, only 67% of the remaining proteins were predicted being targeted to the chloroplast. Out of the 318 identified proteins in the soluble fraction, 24 (or 7.5%) contained a predicted SP for the ER (Kleffmann et al., 2004). In another recent study where data from all available plastid proteomic studies was compiled, about 30% of the identified proteins did not possess canonical chloroplast TP (Armbruster et al., 2009). These authors concluded that at least two types of non-canonical chloroplast proteins exist: the inner-envelope type and the ER-dependent type.

### 2.4 Manuscript IV

The specific function of CAH1 in the chloroplast stroma of Arabidopsis is at present unknown. For a long time it was assumed that the chloroplastic $\beta$-CA1, representing up to 5% of the soluble protein in the stroma, could play an important role in optimizing CO$_2$ availability for Rubisco in C3 plants (Badger and Price, 1994). However, several studies using antisense strategies have shown that specific reduction of the expression of $\beta$-CA1 do not cause obvious growth phenotype and had only minor effects on mesophyll CO$_2$ transfer and photosynthesis (Majeau et al., 1994; Price et al., 1994). In another recent study, photosynthetic measurements on a few days old seedlings showed decreased CO$_2$ uptake rates in mutants lacking the $\beta$-CA1 protein, indicating that the protein might play a role in photosynthesis in the early development
of Arabidopsis (Ferreira et al., 2008). Interestingly, seedlings from this mutant that were able to further establish, developed normally and showed no growth phenotype although a strong reduction in CA activity was confirmed. These results suggest that some other CA activity should be responsible for the CO\textsubscript{2} supply at the active site of Rubisco overcoming the low rates at which the uncatalysed interconversion between CO\textsubscript{2} and bicarbonate takes places. The data presented along this thesis clearly established the occurrence of CAH1 in the chloroplast stroma of Arabidopsis, pointing to this protein as a good candidate for this role.

2.4.1 Starch content and CO\textsubscript{2} exchange rates are reduced in CAH1 mutants

Bioinformatic analysis of gene expression, sequence and phenotypic data from Arabidopsis cluster CAH1 together with a group of genes involved in primary chloroplast function, such as Calvin cycle, carbohydrate metabolism, chlorophyll synthesis, among others. Analyses of developed Arabidopsis plants with disrupted CAH1 gene expression revealed a distinct reduction in photosynthetic capacity and starch accumulation (Manuscript IV, Table 1). While no phenotype could be noticed in 4-week old mutant \(\beta\)-CA1 plants (Ferreira et al., 2008), we observed a decrease of up to 40% in CO\textsubscript{2} uptake rates in the CAH1 mutant plants and a concomitant reduction in growth rates and starch accumulation, indicating an important role of CAH1 in the basal function of the plant (Manuscript IV, Table 1 and Figure 3). In agreement with the photosynthesis and starch, the mutant plants showed an alteration of the levels of soluble carbohydrates (Manuscript IV, Figure 4). The most striking differences were observed for carbohydrates related with the chloroplast function, inositol and raffinose. These data point to CAH1 as being an important player in the photosynthetic performance of the plant.

Complementation studies with the HA-tagged protein from manuscript III showed that the N-terminally tagged CAH1 was able to fully restore wt levels of starch in 3-week old plants (Manuscript IV, Figure 5 and 7). Interestingly, complementation with the corresponding C-terminally tagged protein had no, or minor, effect in restoring the mutant phenotype. Why is the C-terminal part of the protein so important for its function? And why does a modest interference, such as presence of the HA-tag, have such deep consequences on functionality of CAH1? Comparison of the C-terminal sequence of CAH1 with the seven other potential \(\alpha\)-CAs in Arabidopsis revealed that CAH1 has a C-terminal extension of about 10 amino acids that is not found in most other \(\alpha\)-CA polypeptides. This part, the C-terminus, is enriched in charged residues (RVDDKETGKNNKKKKPN, last 15 amino acids of CAH1 where basic (R and K) and acidic (D and E) are in bold). The results obtained in manuscript IV show that the C-terminal tagged version of CAH1 was poorly targeted to the chloroplast in Col-0 suspension cells and mainly appeared
associated with the ER and, to a minor extent, with the Golgi apparatus (Manuscript IV, Figure 5). In addition, the tagging of the C-terminus was not only disturbing a signal potentially involved in transport of the protein to the chloroplast and causing its retention in the ER, but also altered the biochemical features of the protein (Manuscript IV, Figure 6 c and d). The data from the complementation analysis of the knockout CAH1 mutants with the two tagged versions revealed relevant information concerning the role of CAH1. The role of CAH1 seems to be directly linked to its localization in close vicinity to the site of CO$_2$ fixation by the Rubisco enzyme and not to the activity of this protein along the route through the endomembrane system to the plastid. From our studies we can conclude that CAH1 is trafficked from the endomembrane system to the chloroplast. In addition, the type of N-linked glycan attached to the protein suggests a pathway for CAH1 to the chloroplast via the Golgi. In the Golgi, CAH1 is sorted from the default secretory pathway and arrives at the chloroplast using, presumably, a novel type of vesicle transport (Figure 2.1). Vesicle fusion at the chloroplast will most likely result in delivery of the polypeptide to the inter-membrane space. Since CAH1 becomes folded in the ER, the protein could not be translocated via the Tic-part of the Toc-Tic system, suggesting that occurrence of some kind of novel and specific transporter at the inner membrane must be involved.

We believe that the C-terminus of CAH1 is, in some way, involved in this targeting, either as a sorting signal at the Golgi or in assisting transport over the inner envelope membrane; and that interference of the C-terminus results in a mistargeted and therefore non-functional enzyme. Interestingly, some bacterial proteins are transported across the cytoplasmic membrane in folded conformation. Transport mediated by this system requires a twin-arginine (RR) consensus, therefore called the Tat-pathway (Twin-arginine translocation). This pathway is also involved in the transport of some proteins over the thylakoid membrane (Muller and Klosgen, 2005). Since the inner envelope membrane of the chloroplast is believed to have a cyanobacterial origin, a similar system could exist for translocation of folded proteins from the inter-membrane space to the stroma, and the C-terminus of CAH1 could be involved in mediating this step.

The inability of the C-terminally tagged protein to complement the CAH1 KO phenotype raises several questions about targeting, regulation and function of CAH1 that will require extensive research to be answered.
The work presented in this thesis shows two aspects that will change our view of intracellular protein trafficking in plants. First of all, it reveals the presence of a direct pathway from the endomembrane system to the chloroplast. Second, it also shows the occurrence of N-glycoproteins in an endosymbiotic organelle such as the chloroplast. So far, targeting of nuclear-encoded polypeptides to the chloroplast interior was believed to occur exclusively via the Toc-Tic system, a translocon complex in the chloroplast envelope where proteins which were translated in the cytosol are transported across the double membrane of the chloroplast in an extended conformation. In this thesis, an additional pathway for N-glycosylated proteins from the endomembrane system to the chloroplast is presented.

The existence of proteins derived from the endomembrane system in the chloroplast had already been proposed, although conclusive evidence describing this phenomenon was missing (Chen et al., 2004). In this work we show that an \( \alpha \)-type carbonic anhydrase (CAH1) in the model plant \emph{Arabidopsis thaliana} is localized to the chloroplast stroma, although it is first directed to the ER where N-linked glycans are acquired. The type of glycan attached to CAH1 suggested that the protein is trafficked through the Golgi apparatus on its way to the chloroplast. Shortly after our publication, other studies describing a similar transport pathway for two different proteins to the chloroplast in rice were published. As with CAH1, these proteins were N-glycosylated in the ER and their trafficking to the chloroplast was sensitive to exposure of an ER-to-Golgi trafficking inhibitor. In contrast to CAH1, both rice proteins presented a different type of N-linked glycan, suggesting a slightly different (possibly direct) route from the ER to the chloroplast. Our work, together with these two reports, indicates that two individual pathways for targeting of N-glycoproteins to the chloroplast might co-exist in plants.

Development of a transient co-expression system enabled us to show that small GTPases involved in canonical ER-to-Golgi vesicle transport were required for trafficking of CAH1 to the chloroplast. A question that arose was why a small set of proteins that are targeted to the chloroplast follow this pathway, while the vast majority of the nuclear-encoded chloroplast proteins are transported through the Toc-Tic complex? We characterized two types of protein modifications taking place in the ER/Golgi: N-glycosylation and disulphide bridge formation. CAH1 was found to be glycosylated at four or five sites and to possess an intra-molecular disulphide bridge. Both types of
protein modifications were required for correct folding of the CAH1 polypeptide, since protein lacking any of these modifications formed aggregates in the ER, unable to be targeted further. In addition, we also showed the importance of these post-translational modifications for producing a functional CA.

Based on these findings we propose that the protein is not functional without endomembrane-associated post-translational modifications. As a consequence, a pathway to the chloroplast from the secretory system (where the protein is modified to be functional) is required. Since the unfolding of the protein, required for crossing the chloroplast envelope through the Toc-Tic, most certainly would result in a CAH1 with impaired structure and function, this alternative pathway should be independent of the Toc-Tic machinery.

The existence of such a pathway for proteins to the chloroplast implies that the activities of plastid N-glycoproteins, such as CAH1, are relevant for the function of the chloroplast. A genetic dissection using Arabidopsis plants with disrupted CAH1 gene expression showed that the activity of this CA plays a pivotal role in the photosynthetic performance of the plant. Mutant plants exhibited lower photosynthetic capacity and reduced starch content. An allele with an especially strong mutant phenotype additionally showed retarded growth and altered levels of soluble carbohydrates. In agreement with the mutant phenotype, bioinformatic analysis indicated that CAH1 is co-expressed with other chloroplast proteins involved in basic functions of the chloroplast such as photosynthesis and synthesis of photosynthetic pigments.

In conclusion, this thesis presents firm evidence for a pathway of proteins from the endomembrane system to the chloroplast in *Arabidopsis thaliana*. We show that trafficking of proteins along this route depends on canonical mechanisms, such as ER-directed translation and ER-to-Golgi vesicle transport. Study of the mechanisms involved in transport of CAH1 to the chloroplast can reveal crucial information for our understanding about chloroplast, and plant cell evolution. Gene expression data, together with the phenotype observed in CAH1 mutant plants, indicate that the CAH1 protein plays a crucial role in the photosynthetic process. Activity measurement of over-expressed CAH1 showed that ER-associated post-translational modifications, such as N-glycosylation and disulphide bridge formation, are not only important for folding and trafficking of the protein, but also for its function, possibly explaining why this pathway (still?) exist in plants.

### 3.1 Future perspectives

Several questions remain to be answered, both regarding the mechanism of this pathway to the chloroplast and the exact function of CAH1. For example,
which signals target CAH1, and other glycoproteins (?), from the endomem-
brane system to the chloroplast? Are there one or several pathways from the
secretory system to the chloroplast? What is the function of the C-terminal
part of the CAH1 protein? How is CAH1 crossing the inner membrane of
the chloroplast envelope? Is there an active transporter for CAH1 across the
membrane? In what way is CAH1 involved in the photosynthetic process?
Which other proteins follow the same route as CAH1? What is the evolution-
ary reason for having two (or more) protein pathways to the chloroplast?

What initially attracted my attention to this project was the emerging poten-
tial of using plants as bio-factories for production of recombinant proteins, and
specifically how this pathway could be used for accumulation of chloroplast-
localized recombinant glycoproteins. Many commercially interesting proteins
require N-linked glycosylation for their functionality, e.g. antibodies. The
chloroplasts (or plastids in general) have been shown to be able to accumu-
late large amounts of foreign proteins without interfering with the function
of the organelle. Until now, the requirement of N-linked glycans has directed
expression and accumulation of the proteins to compartments such as the ER
or the vacuole, or to be secreted. The possibility of combining expression of
N-glycosylated polypeptides in plants with targeting to and accumulation in
plastids emerged to me as an appealing alternative to existing methods. What
is clear to me now, after 5 years of (hard?) work, is that ideas which appear
to be straightforward to implement are not always that easy to carry out in
practice, and that intense continued research in this field will be needed.
Acknowledgement

Slim For believing in me from the beginning and for always supporting me in whatever crazy idea/experiment I have had, for making me continue when my belief in the project has been low, for letting me do part of the work outside Umeå, for all late-night changes of antibody solutions...

Arsenio For not taking your hand away from my project even when moving 4000 km away, for letting me come to Madrid and do part of the work there, for all the help with the thesis and manuscripts the last crazy days before Christmas. It would not have been possible without you.

Laszlo Bako For helping me maintaining my dear cell suspension cultures during my frequent trips to Spain, for knowing everything about biochemistry and for becoming more than a colleague.

Markus Grebe For always being interested in my project, for bringing a lot of new knowledge to UPSC and for all visits to my office.

Daniel Ziggy, Diamyd-Danne, Herr Toffel...Kärt barn har många namn. Tack för alla luncher nere på stan eller IKSU, alla fika, för allt ovårderligt börssnack och för att alltid lyckas inkludera USA’s handelsunderskott i varje diskussion (oavsett tema)...

Max För att vara en trogen kompis och hockeysupporter, för din mycket speciella humor, för alla stunder på innebandyplanen. Även om min publikationslista aldrig kommer att vara lika lång som din, så har jag i alla fall spikat Organisk kemi B...Så det så.

Rolleboll För alla konstiga mail som har gjort mig glad, för att kunna Bilbao’s bakgator utan och innan, för alla Guggenheim-besök. Dock besviken på att du sticker från Spanien nu, förväntar mig att du flyttar tillbaka snarast...

Cristian For being a really really nice friend, for always being optimistic, for all squash games and discussions in the sauna, for showing me Madrid.

Maribel and Markus For being so extremely nice. Please move down soon.

Andreas För att alltid ställa upp, för att alltid kunna besvara datorrelaterade frågor, för att visa mig krångligast möjliga dataprogram för att sätta samman denna avhandling.

Claes För vår goda vänskap sedan 1-års ålder, för all hjälp när jag flyttat hit och dit och för att alltid låtit mig bo hos dig närhelst det behövts. Förväntar mig att du kommer och besöker mig ofta.
Martin Frenkel For din tröstsockerkaka när AIK lyckades att misslyckas med att ta klivet upp i elitserien våren 2004.

Juande For your laughter and for always losing against me in squash.

Sergey and Grit For BBQs and fikas. It is a shame that I have to leave just after getting to know you. You are welcome to Spain whenever . . .

Innebandygänget Vem ska nu hacka på Daniel’s smalben?

Tatiana For standing my taste of music, for always being nice and helpful in the office. We all hope, want and pray that you will be back soon.

Adele Many thanks for reading my thesis pre-printing and improving my Swenglish, for not being as dangerous as you look and for being a really nice person in the lab.

All past Slim group members Linda, Dan . . . that has made the lab a good place to work.

All past UPSC members that has made UPSC a fun place to be. Steve, Aurora, Olivier, Frank . . .

All UPSC It has been fun to be part of UPSC these years and I will miss you all.

All people outside UPSC that has helped me and been part of my life during these years.

All my new friends in Madrid Flor for helping me to feel at home in the lab, Amaya and Andrea for collaborations, yoghurt-Ismael for having 12.30 fika with me, Lars for proving to me that also Spanish people can listen to real music, and many more . . .

Marisa and family For letting me stay at your place in Madrid and making me have a second family.

Jose and Chelo For accepting me although I don’t know the difference between ”Nieve, Nuevo and Nueve” or ”Silla and Casilla”. I will improve, I promise!

Familjen in Rönninge För alla besök som jag kunnat göra i samband med Spanienvistelser, trots att det måste varit jobbigt.

Jonas För att vara en bra bror som mer än gärna korrekturläser under helger.

Mina föräldrar För att vara så förstående.

Cris FOR EVERYTHING!!
References


Stefano, G., Renne, L., Chatre, L., Hanton, S. L., Moreau, P., Hawes, C., and Brandizzi, F. (2006): In tobacco leaf epidermal cells, the integrity of protein export from the endoplasmic reticulum and of ER export sites depends on active COPI machinery. The Plant Journal 46, 95-110.


