The chloroplast lumen
- New insights into thiol redox regulation and functions of lumenal proteins

Michael Hall
“Height has nothing to do with it, it is your strength that counts.”
- Lynn Hill, the Nose, Yosemite
Table of Contents

Table of Contents i
List of original publications ii
Abstract iv
Abbreviations v
Introduction 1
  Oxygenic photosynthesis 1
  The leaf chloroplast - origin and architecture 2
  The envelope membrane 3
  The soluble stroma compartment 5
  The thylakoid membrane and photosynthetic electron transport 5
  The ferredoxin thioredoxin system of oxygenic photosynthesis 7
  The thylakoid lumen 9
  Protein import into the lumen 9
  Overview of the lumen proteome 11
  Functions of luminal proteins 12
Aim of this thesis 19
  General aim 19
  Specific aims 19
Results and discussion 21
  Diurnal regulation of the lumen proteome (Paper I) 21
  Identification of thioredoxin targets in the chloroplast lumen (Paper II) 26
  Localization studies of chloroplastic thioredoxins (Paper III) 32
  Characterization of the pentapeptide repeat proteins (Paper IV) 36
  Purification and crystallization of PPD6, a PsbP-domain protein (Paper V) 40
Summary and conclusions 43
Future perspectives 51
Acknowledgements 52
References 55
List of original publications

This thesis is based on the following publications, referred to by their roman numerals throughout the text.


Publications not included in this thesis


# These authors contributed equally to this work.

Abstract

In higher plants oxygenic photosynthesis primarily takes place in the chloroplasts of leaves. Within the chloroplasts is an intricate membrane system, the thylakoid membrane, which is the site of light harvesting and photosynthetic electron transport. Enclosed by this membrane is the lumen space, which initially was believed to only contain a few proteins, but now is known to house a distinct set of >50 proteins, many for which there is still no proposed function. The work presented in this thesis is focused on understanding the functions of the proteins in the lumen space. Using proteomic methods, we investigated first the regulation of lumenal proteins by light and secondly by dithiol-disulphide exchange, mediated by the disulphide reductase protein thioredoxin. We furthermore performed structural and functional studies of the lumenal pentapeptide repeat proteins and of the PsbP-domain protein PPD6.

When studying the diurnal expression pattern of the lumen proteins, using difference gel electrophoresis, we observed an increased abundance of fifteen lumen protein in light-adapted Arabidopsis thaliana plants. Among these proteins were subunits of the oxygen evolving complex, plastocyanin and proteins of unknown function. In our analysis of putative luminal targets of thioredoxin, we identified nineteen proteins, constituting more than 40 % of the lumen proteins observable by our methods. A subset of these putative target proteins were selected for further studies, including structure determination by x-ray crystallography. The crystal structure of the pentapeptide repeat protein TL15 was solved to 1.3 Å resolution and further biochemical characterization suggested that it may function as a novel type of redox regulated molecular chaperone in the lumen. PPD6, a member of the PsbP-family of proteins, which is unique in that it possesses a conserved disulphide bond not found in any other PsbP-family protein, was also expressed, purified and crystallized. A preliminary x-ray analysis suggests that PPD6 exists as a dimer in the crystalline state and binds zinc ions.

The high representation of targets of thioredoxin among the lumen proteins, along with the characterization of the pentapeptide repeat protein family, implies that dithiol-disulphide exchange reactions play an important role in the thylakoid lumen of higher plants, regulating processes such as photoprotection, protein turnover and protein folding.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>APR</td>
<td>adenylylsulphate reductase</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>Ctp</td>
<td>carboxyl-terminal processing peptidase</td>
</tr>
<tr>
<td>DIGE</td>
<td>difference gel electrophoresis</td>
</tr>
<tr>
<td>DMA</td>
<td>dimethylacrylamide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>Fd</td>
<td>ferredoxin</td>
</tr>
<tr>
<td>FNR</td>
<td>ferredoxin-NADP+ oxidoreductase</td>
</tr>
<tr>
<td>FTR</td>
<td>ferredoxin-thioredoxin reductase</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>HCF</td>
<td>high chlorophyll fluorescence</td>
</tr>
<tr>
<td>IAM</td>
<td>iodoacetamide</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>LHC</td>
<td>light harvesting complex</td>
</tr>
<tr>
<td>MBB</td>
<td>monobromobimane</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>NADPH</td>
<td>nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NDH</td>
<td>NAD(P)H dehydrogenase</td>
</tr>
<tr>
<td>NEM</td>
<td>N-ethylmaleimide</td>
</tr>
<tr>
<td>OEC</td>
<td>oxygen evolving complex</td>
</tr>
<tr>
<td>PLAS</td>
<td>plastocyanin</td>
</tr>
<tr>
<td>PPD</td>
<td>PsbP-domain</td>
</tr>
<tr>
<td>PPL</td>
<td>PsbP-like</td>
</tr>
<tr>
<td>PPR</td>
<td>Pentapeptide repeat protein</td>
</tr>
<tr>
<td>PQL</td>
<td>PsbQ-like</td>
</tr>
<tr>
<td>PSI</td>
<td>photosystem I</td>
</tr>
<tr>
<td>PSII</td>
<td>photosystem II</td>
</tr>
<tr>
<td>RuBisCo</td>
<td>ribulose-1,5-bisphosphate carboxylase/oxygenase</td>
</tr>
<tr>
<td>SEC</td>
<td>size-exclusion chromatography</td>
</tr>
<tr>
<td>SRP</td>
<td>signal recognition particle</td>
</tr>
<tr>
<td>TIC</td>
<td>translocon at the inner envelope membrane of chloroplasts</td>
</tr>
<tr>
<td>TOC</td>
<td>translocon at the outer envelope membrane of chloroplasts</td>
</tr>
<tr>
<td>TPP</td>
<td>thylakoid processing protease</td>
</tr>
<tr>
<td>Trx</td>
<td>thioredoxin</td>
</tr>
<tr>
<td>VDE</td>
<td>violaxanthin de-epoxidase</td>
</tr>
</tbody>
</table>
**Introduction**

**Oxygenic photosynthesis**

Intricately linked to all life on our planet, the origin and evolution of photosynthesis is a research topic of great interest. While a combination of genetical, biochemical and geophysical information is increasing our understanding of photosynthesis, many aspects still remain unknown, with several different models being proposed (Hohmann-Marriott and Blankenship, 2011). It is generally accepted that more than 3.3 billion years ago, the ancestors of cyanobacteria gained the ability to use light energy from the sun together with hydrogen sulphide or iron to produce chemical energy, in a process called anoxygenic photosynthesis. This process was performed using a protein complex referred to as a type I photosystem, the ancestor of photosystem I found in cyanobacteria, algae and plants. According to one model, the single type I photosystem became duplicated and the two separate photosystems diverged to gain distinct functions (Mulkidjanian et al., 2006; Allen and Martin, 2007). Before evolving the ability to split water the ancestral procyanobacterium possibly expressed one type of photosystem at a time, depending on the demands, with the help of a regulatory switch. A mutation in this switch at the correct time and environmental setting then led to the simultaneous expression of the two photosystems (Allen and Martin, 2007). With the addition of a catalyst that oxidizes water, for example manganese atoms, the two photosystems would become complementary, spanning a larger redox potential gap, enabling linear electron flow and finally leading to the emergence of oxygenic photosynthesis in the first true cyanobacteria, around 3.4-2.3 billion years ago (Bekker et al., 2004; Allen and Martin, 2007). Now utilizing water, one of the most abundant molecules on earth, and working in concert with the fixation of atmospheric carbon dioxide via the Benson-Calvin cycle, the process of oxygenic photosynthesis was able to use light together with water and CO₂ to produce molecular oxygen and carbohydrates. The overall chemical reaction of oxygenic photosynthesis can be summarized as below, where it should be noted that the oxygen derives from water and not CO₂.

\[ 6\text{CO}_2 + 6\text{H}_2\text{O} + \text{light} \rightarrow \text{C}_6\text{H}_{12}\text{O}_6 + 6\text{O}_2 \]

This intricate process was passed on from cyanobacteria to algae and plants, laying the foundation for the earth's biosphere as we know it today. Oxygenic photosynthesis accounts for about 98% of the oxygen in our atmosphere, while the remaining 1-2% is formed by cleavage of water molecules by ultraviolet radiation. Although still under debate, it is estimated that 30-50% of this oxygen is produced by algae and cyanobacteria in the oceans, while the remaining oxygen is produced by land plants.
The leaf chloroplast - origin and architecture

In plants photosynthesis takes place in mesophyll and bundle sheath cells, the major components of leaves. More specifically it takes place in a cellular organelle called the chloroplast, of which there are ~120 in a typical mature *Arabidopsis thaliana* leaf mesophyll cell (Buchanan et al., 2002). It is within the chloroplast that we find the chlorophyll molecules which absorb the light energy from the sun and make our planet green. It is also here that the large photosynthetic protein complexes controlling oxygen evolution and CO$_2$-fixation are located. The intricate architecture and precise function of these protein complexes and pigment molecules will be discussed in more detail further on. Chloroplasts are considered to have originated from cyanobacteria through a process known as endosymbiosis, first described by Konstantin Mereschkowski in 1905. During endosymbiosis a cyanobacterial cell was engulfed by a eukaryotic cell and the major part of the cyanobacterial genome was transferred to the nuclear genome of the eukaryotic cell over evolutionary time. While cyanobacteria typically have genomes encoding several thousand proteins (Beck et al., 2012), the chloroplast genomes usually encode only 60-200 proteins, constituting only ~5-10% of the original cyanobacterial genome (Martin and Herrmann, 1998). While the eukaryotic cell inherited many useful functions from the cyanobacteria, such as photosynthesis, 70S ribosomes and cell division proteins, it also needed adaption and invention of other processes (Martin et al., 2002). An important process which required adaption was a transport system for those proteins previously encoded by the cyanobacterial genome but now encoded in the eukaryotic cell nucleus (Heins and Soll, 1998; Martin et al., 2002). Arabidopsis chloroplasts have been predicted to contain 2090 proteins, similar in number to cyanobacterial cells, out of which about half have been experimentally identified in proteomic studies and can be accessed in databases such as plprot, SUBA and PPDB (Kleffmann et al., 2006; Heazlewood et al., 2007; Zybaiov et al., 2008).

Leaf chloroplasts commonly have a hemispherical shape and measure approximately 5-10 µm in the long dimension and 3-4 µm in the short dimension. The chloroplast is enclosed by two membranes, the outer envelope membrane and the inner envelope membrane, which together enclose the envelope membrane space. According to the endosymbiotic theory, the inner envelope corresponds to the plasma membrane of the cyanobacterial cell, while the outer membrane corresponds to a vesicle of the eukaryotic host’s plasma membrane, formed as the cyanobacterial cell was engulfed. The soluble compartment within the chloroplast is known as the stroma and it contains a high concentration of proteins of diverse function, most prominently important metabolic enzymes. Throughout the chloroplast is a continuous internal membrane system, the thylakoid membrane, consisting of stacked cylindrical membrane regions called grana and
connecting regions termed stroma lamellae (Austin and Staehelin, 2011). The thylakoid membrane is where chlorophyll molecules are bound to large protein complexes, the photosystems, and where the energy from the sunlight is absorbed and converted to primary forms of chemical energy in the form of ATP and NADPH. Enclosed by the thylakoid membrane is a narrow soluble space called the chloroplast- or thylakoid lumen. Originally it was believed that the lumen space was more or less devoid of proteins and contained only a few soluble components of the two photosystems, plastocyanin and salts. However, it was shown in 1998 that this compartment indeed contained a distinct set of proteins (Kieselbach et al., 1998). The thylakoid lumen has been the main focus of the work in this thesis and it will be revisited in great detail in the coming chapters.

The envelope membrane

Across the chloroplast envelope membrane an extensive trafficking of photosynthetic products, metabolic intermediates and proteins takes place. The outer envelope membrane was first considered to be a non-specific molecular sieve, allowing small molecules such as water, ions and metabolites to pass freely into the intermembrane space. The discovery of
proteins which form solute channels with distinct specificities for different metabolites has however shown that things are more complicated, and that the outer membrane may play a much more dynamic role (Duy et al., 2007). Besides being involved in control of metabolic fluxes between the plastid and cytosol, the outer envelope plays an important role in protein import to the chloroplast, and is also metabolically active, for example playing a role in membrane lipid production and fatty acid metabolism (Breuers et al., 2011). While the inner envelope membrane is believed to be freely permeable to small uncharged molecules such as O$_2$, CO$_2$ and NH$_3$, this view has recently been challenged with the discovery of inner envelope aquaporins. Uehlein et al. showed that RNA interference lines of the AQP1 aquaporin had greatly reduced permeability of CO$_2$ at the inner envelope membrane. This has called into question both the functions of aquaporins as well as the concept of free diffusion of gases such as CO$_2$ across biological membranes in general (Uehlein et al., 2008). Larger molecules, such as the many metabolites passing the membrane, however, certainly require specific protein transporters in order to pass the membrane. A proteomic mapping study, specifically focused on the envelope fraction, identified 298 proteins located in the envelope proteome (Ferro et al., 2010). Of these proteins 19% were functionally classified as being involved in metabolism, 24% were represented by transporters, 8% by chaperones/proteases and 10% by protein targeting related proteins. In other words, these main groups comprise more than 60% of the envelope proteome. Very interesting are of course also the remaining proteins of the envelope membrane system, out of which 21% have unknown function and the remaining 18% are involved in processes such as redox regulation, stromal translation, stress response, signaling and other processes.

As mentioned previously, one very important function of the envelope membrane is to facilitate the import of stromal, thylakoid and luminal proteins into the chloroplast. This takes place via a common import apparatus, namely the TOC/TIC (translocon at the outer/inner envelope membrane of chloroplasts) complexes, located at the outer- (TOC) and inner envelope (TIC) membranes (Soll and Schleiff, 2004). In the first stage of import, a chloroplast targeted pre-protein is synthesized on an 80S ribosome in the cytosol. The pre-protein then binds reversibly with receptor components of the TOC complex. GTP in the cytosol and low concentrations of ATP in the intermembrane space are then required for the pre-protein to be transported through the outer envelope membrane. The pre-protein immediately comes in contact with the TIC complex and is translocated across the inner envelope, pulled through by the stromal HSP93 chaperone, in a process catalyzed by ATP in the stroma (Flores-Perez and Jarvis, 2012).
The soluble stroma compartment

While the light harvesting and photosynthetic electron transport reactions take place in the thylakoid membrane, the second component of the photosynthetic process, the fixation of atmospheric carbon into carbohydrates, takes place in the soluble stroma compartment of the chloroplast. Central in this process is the Benson-Calvin cycle, where carbon dioxide, NADPH and ATP are used to produce carbohydrates in the form of triose phosphates, by a set of stromal enzymes. The Benson-Calvin cycle is linked to the light reactions of photosynthesis via a regulatory system, the ferredoxin-thioredoxin system, which will be described in more detail in a coming section. For the model plant species *Arabidopsis thaliana* it has been estimated that 76 % of the total protein content in the stroma constitutes metabolic proteins involved in the Benson-Calvin cycle, glycolysis and the oxidative pentose phosphate pathway (Peltier et al., 2006). Principle among the enzymes required for CO₂ fixation is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which catalyzes the carboxylation of ribulose-1,5-bisphosphate using CO₂, in other words the first step in the fixing of atmospheric CO₂. In contrast to most textbook figures and the figure above (Figure 1), where the stroma looks mostly empty besides some starch and chloroplast DNA, the stroma has a very high concentration of protein. The stroma has a protein concentration around 400 mg/ml, out of which RuBisCO alone has a concentration of 250 mg/ml (Yokota and Canvin, 1985; Yabuta et al., 2008). This means that RuBisCo constitutes 60-70 % of the protein content in the stroma and around 30-50 % of the soluble protein content in leaves. Even though it is dominated to a large extent by metabolic enzymes, the stromal compartment houses many other proteins of lower abundance which function in a myriad of pathways.

The thylakoid membrane and photosynthetic electron transport

The thylakoid membrane is mainly dominated by four differentially distributed protein complexes, which constitute the photosynthetic electron transport chain. The photosystem I (PSI) and ATP-synthase (ATPase) complexes are located exclusively to the non-stacked stroma regions and the unappressed grana regions of the thylakoid membrane. The photosystem II (PSII) complex is however predominantly located in the stacked grana regions, while the cytochrome *b₅f* (Cyt *b₅f*) complex is heterogenously distributed throughout the membrane (Albertsson, 2001).

In linear photosynthetic electron transport the two principal products O₂ and NADPH are produced (Figure 2). In the first step, light energy is absorbed by chlorophyll molecules bound to light harvesting complexes (LHCs), and is further funneled to the reaction centres of PSI and PSII by resonance energy transfer. In the PSII reaction center this leads to the excitation of an electron
in a specific chlorophyll $a$ molecule referred to as P680. The excited electron is then rapidly transferred to the primary electron acceptor pheophytin (Phe) and subsequently to the plastoquinones $Q_A$ and $Q_B$ on the stromal side of the thylakoid membrane. The oxidized P680$^+$ is reduced by a redox-active tyrosine residue (Y$_Z$) in the reaction center D$_1$ protein, which in turn has extracted the electron via the manganese cluster from the oxidation of a water molecule. After $Q_B$ has accepted two electrons it acquires two protons from the stroma, transforming it to a hydrophobic plastoquinol molecule (PQH$_2$), which diffuses in the membrane. The $Q_B$ site on the PSII reaction center protein is filled with a new plastoquinone molecule from the pool of free quinones diffusing in the membrane. In the next step the Cyt $b_6f$ complex then mediates the transfer of the electrons from PQH$_2$ to the luminal plastocyanin protein, with the two protons being released in the lumen. In PSI, light energy excites a chlorophyll $a$ molecule in P700, which transfers an electron to a primary electron acceptor, the chlorophyll $a$ molecule A$_0$. The oxidized P700$^+$ is then reduced back by plastocyanin. From A$_0$, the electron is transferred via a phylloquinone (A$_1$) and several iron-sulphur centres to the ferredoxin protein on the stromal side of the thylakoid membrane. Finally the electron is used to reduce NADP$^+$ to NADPH by the ferredoxin-NADP$^+$ oxidoreductase (FNR). Due to the release of protons on the lumen side during the electron transport, a proton gradient ($\Delta$pH) is generated across the thylakoid membrane. The accumulation of protons in the lumen is ultimately used to drive the synthesis of ATP from ADP in the chloroplast stroma by the fourth major complex in the thylakoid membrane, the ATP synthase.

**Figure 2.** Linear photosynthetic electron transport. Abbreviations are explained in the text and abbreviation list. The protein representations were made from the previously determined crystal structures obtained from the Protein data bank (Pdb): 1S5L (Photosystem II from T. elongatus), 1VF5 (Cytochrome b$_6f$, M laminosus), 1AG6 (Plastocyanin, S. oleracea), 1JB0 (Photosystem I, S. elongatus), 1A70 (Fd, S. oleracea), 2VNH (FNR, R. capsulatus) and 1E79/2W5J (ATP synthase, B. taurus/S. oleracea).
An interesting aspect, with the roles of proteins in the lumenal compartment in consideration, is the organization of individual stacked thylakoid membranes in the grana regions. Kirchhoff et al. estimated that 80% of the total area of a single grana disc is occupied by protein (predominantly PSII and LHC trimers) and only 20% by lipids, making grana thylakoids and hence the lumen compartment very crowded (Kirchhoff et al., 2004). This is especially important when considering the required lateral movement of certain proteins, such as plastocyanin. However in a recent study the same group showed that the thylakoid lumen expands in the light and contracts in the dark. According to their calculations the distance between opposing thylakoid membranes was 47 ± 8 Å for dark adapted leaves and 92 ± 6 Å for light adapted leaves, a more or less doubled distance (Figure 3)(Kirchhoff et al., 2011). Such an increase in the lumen volume would facilitate easier lateral protein movement in the lumen in conditions when the need is increased. But even in the light the lumen is a narrow space. For example the oxygen evolving complex (OEC) of PSII protrudes 40 Å into the lumen, meaning that two face-to-face PSII complexes would almost be in contact with each other, spanning the lumen space. It importantly also suggests that the majority of lumen proteins most certainly are in contact with and interact with proteins and complexes in the thylakoid membrane.

The ferredoxin thioredoxin system of oxygenic photosynthesis

An important regulatory link between the light reactions of photosynthesis and metabolic processes in the chloroplast was established with the discovery that light driven covalent modifications of cysteine amino acids in Benson-Calvin cycle enzymes could regulate their activity. In the light, reduced ferredoxin reacts with and reduces a disulphide bond in oxidized thioredoxin, catalyzed by the enzyme ferredoxin-thioredoxin reductase (FTR)(Wolosiuk and Buchanan, 1977a). The reduced thioredoxin protein can
then in turn reduce disulphide bonds in target proteins, which may induce conformational changes and modulate their enzymatic activities, either activating or deactivating them. This system is known as the ferredoxin thioredoxin system and was initially shown to activate NADP⁺-malate dehydrogenase and several important Benson-Calvin cycle enzymes in the light, including fructose-1,6-bisphosphatase, sedoheptulose-1,7-bisphosphatase, phosphoribulokinase and NADP⁺-glyceraldehyde-3-phosphate dehydrogenase (Wolosiuk et al., 1977b; Breazeale et al., 1978; Wolosiuk and Buchanan, 1978a, 1978b; Buchanan, 1980). In this manner biosynthetic enzymes in the stroma are activated by reduction in the light (S-S -> 2SH), while catabolic enzymes, which are active in the disulphide state (S-S), instead become de-activated in the light. Since this landmark discovery, thioredoxin mediated regulation has been found to occur in a multitude of important cellular pathways in bacteria, animals and plants, facilitated greatly by the advent of proteomic methods for target identification. Thioredoxin targets have been found in all compartments of the chloroplast, from the envelope membrane to the thylakoid lumen, representing among many other processes: photosynthesis, the Benson-Calvin cycle, starch and glycogen synthesis, sulfur and nitrogen metabolism, oxidative stress and protein import (Buchanan and Balmer, 2005; Lindahl and Kieselbach, 2009; Balsera et al., 2010). Using chloroplast preparations from Arabidopsis, spinach, poplar and Chlamydomonas, more than 100 in vitro targets of thioredoxin and glutaredoxin (another thiol-dependent redox active enzyme) have been identified to date (Motohashi et al., 2001; Balmer et al., 2003; Lemarie et al., 2004; Marchand et al., 2004; Balmer et al., 2006; Marchand et al., 2006; Motohashi and Hisabori, 2006; Bartsch et al., 2008). In cyanobacteria, the probable ancestors of plant plastids, roughly 80 target proteins have been identified, representing essentially the same processes as those found to be thioredoxin-linked in chloroplasts (Lindahl and Florencio, 2003, 2004; Perez-Perez et al., 2006; Mata-Cabana et al., 2007; Lindahl and Kieselbach, 2009). The complexity of the thioredoxin system is increased by the presence of several different types of thioredoxins.
in the chloroplast. 19 genes encoding different thioredoxin and putative thioredoxin proteins have been identified in the genome of Arabidopsis (Meyer et al., 2005). The classical m-type and f-type thioredoxins consist of four and two members respectively while the more recently discovered y-type and x-type have two and a single member respectively. The remaining ten genes encode atypical thioredoxins which differ slightly in their active site or consist of multiple domains. Also a recent addition to the thioredoxin family was made with the discovery of a new z-type, which consists of a single member in Arabidopsis (Arsova et al., 2010). The sub-chloroplastic localization of these proteins has been examined in Paper III of this thesis and they will therefore be described in more detail in the section summarizing that work. While a lot of work has been performed to understand the ferredoxin/thioredoxin system, including the structure determination of its components and several target proteins, little is still understood about the re-oxidation of target proteins back to their disulphide form, which would complete the regulatory cycle.

The thylakoid lumen

Protein import into the lumen

Soluble thylakoid lumen proteins are first imported into the chloroplast by the TOC/TIC complex and are then directed into the lumen, through either the Tat (twin-arginine translocase) pathway, which is \( \Delta \text{pH} \)-dependent, or the Sec (secretory) pathway, which requires ATP (Figure 5)(Jarvis and Robinson, 2004; Aldridge et al., 2009). Lumenal proteins are synthesized in the cytosol with an N-terminal bipartite transit peptide (none of the known lumen proteins are encoded in the plastid genome). The first part of the signal peptide directs the precursor protein to the chloroplast, via the TOC/TIC complex, and is then cleaved off in the stroma by a stromal processing peptidase. The second part of the signal peptide is exposed, directing the protein further into the lumen via the Tat or Sec pathway, where it is finally cleaved off, in this case by the thylakoid processing peptidase (TPP). Proteins targeted to the thylakoid membrane are inserted into the membrane using two pathways. The first is mediated by the chloroplast signal recognition particle (cpSRP), in a pathway similar to that used by bacteria for insertion of almost all membrane proteins of the inner membrane. While the SRP pathway is used for the highly abundant light harvesting chlorophyll binding proteins, most other thylakoid proteins use a second import pathway, so far referred to as the ‘spontaneous’ pathway, which is completely different and does not use SRP or any other known targeting apparatus.

The lumenal signal peptides are generally similar for import via both the Tat and Sec pathways. The N-terminal is characterized by a serine and threonine
rich region for the passage across the envelope membrane, and they then contain a hydrophobic core region, approximately 15-20 amino acids in length, before the TPP processing site. The processing site for both the Tat and Sec pathways can be described by the consensus motif Ala-X-Ala. A distinct element of the Tat pathway is a twin-arginine motif (RR), located just prior to the hydrophobic core region. Signal peptides for the Sec pathway in contrast do not have this motif but instead typically have a charged lysine residue next to the N-terminus of the hydrophobic core (Albiniak et al., 2012). An important difference between the Tat and Sec pathways is that while proteins imported by the Sec pathway do so in an unfolded state, proteins passing through the Tat complex can do so in a fully folded state (Hynds et al., 1998; Marques et al., 2004). In the original mapping of the Arabidopsis lumen proteome, Schubert et al. (2002) found a quite even distribution of Tat and Sec targeted lumen proteins, with 19 containing the twin arginine motif, targeting them via the Tat translocon, and 16 having Sec pathway transit peptides.

Figure 5. Protein import to the thylakoid membrane and thylakoid lumen. Proteins are translated in the cytoplasm with an N-terminal transit peptide (green). Following translocation into the chloroplast via the TOC/TIC complexes, the transit peptide is cleaved. Proteins targeted to the lumen are further imported in a folded or unfolded state via the Tat pathway (red) or in an unfolded state via the Sec pathway (yellow). Integral thylakoid proteins are inserted into the membrane using either the SRP mediated or the so called ‘spontaneous’ pathway. Adapted from Jarvis and Robinson, 2004.
Overview of the lumen proteome

The chloroplast lumen plays a central role in the formation of the thylakoid proton gradient that drives the synthesis of ATP, making it essential for oxygenic photosynthesis. For many years the presence of some lumenal proteins such as plastocyanin and the extrinsic subunits of Photosystem II, had been known, but the lumen space was not considered to be a compartment where networks of other biochemical pathways were taking place. In 1998 the first systematic study of chloroplast lumen proteins in contrast showed that the lumen in fact contained a considerable number of proteins of unknown function (Kieselbach et al., 1998). Following this discovery several research groups began to apply proteomic methods in order to map the proteome of the lumen (Kieselbach et al., 2000; Peltier et al., 2000; Bricker et al., 2001). These studies were however limited by lack of complete sequencing information, and it wasn’t until the sequencing of the Arabidopsis thaliana genome, that the first proper characterization of the lumen proteome could be performed (Schubert et al., 2002). Using a 2D-gel electrophoresis approach, with mass spectrometry and microsequencing for protein identification, Schubert and co-workers could identify 36 luminal proteins. Using this as a basis for a genome-wide prediction they further predicted that the lumenal proteome of Arabidopsis consisted of around 80 proteins. In a similar study, Peltier et al. (2002) also identified 31 luminal proteins and predicted that the Arabidopsis lumen could contain up to 200 proteins. In a review by Kieselbach and Schröder (2003), the knowledge of the lumen proteome at that time was extensively described. During the last ten years, additional proteomic studies (Peltier et al., 2004; Goulas et al., 2006; Zybailov et al., 2008) and functional studies of individual proteins (Gupta et al., 2002a; Petersson et al., 2006) have led to the identification of more soluble lumen proteins, bringing the total number of experimentally verified thylakoid lumen proteins to 53. This is still far from the predicted number of lumen proteins and suggests that many so far undiscovered lumen proteins are of low abundance, or may only be expressed under certain specific conditions. An example of such a protein is the product of the AtCHL gene, a lipocalin domain protein, which rapidly accumulates in the lumen upon dehydration stress, preventing oxidative stress by lipid peroxidation (Levesque-Tremblay et al., 2009). A summary of the known Arabidopsis chloroplast lumen proteome is shown in Figure 6. The proteome of the thylakoid lumen is unique in the sense that it contains no known metabolic enzymes or signaling proteins and it is generally not as diverse as other compartments. Supporting this view, a preliminary metabolomic GC/MS analysis of the lumen fraction used for proteomic studies, did not reveal any metabolites in the lumen (Lindén and Hall, unpublished). It should be noted though, that the lumen isolation method was developed for proteome extraction and may not be optimal for the study of metabolites. The lumen is rather comprised mainly of only a few different protein families, the most prominent being the immunophilin family, with eleven
proteins, the PsbP-like/PsbP-domain family represented by nine proteins, and the pentapeptide repeat family, of which three of the four family members encoded in the Arabidopsis genome are located in the lumen. A comprehensive summary of the currently known lumen proteome, including proposed functions of characterized proteins, import pathways used and references to the papers in this thesis where they are implicated, is found in Table 1 on pages 14-15.

Functions of luminal proteins

The following section will give an overview of the known and proposed functions of luminal proteins, which have been characterized to date. Two prominent protein families, the pentapeptide repeat proteins and the PsbP-like/domain proteins have been a focus of this thesis and will not be described here, but will rather be discussed in detail in the next chapter.

Extrinsic Photosystem II subunits and assembly factors - The PsbO, PsbP and PsbQ proteins are attached to the lumenal side of PSII, constituting what is referred to as the oxygen evolving complex (OEC) in higher plants. They are among the most highly abundant proteins in the lumen, and have been shown to both constitute active OECs, as well as exist as a free soluble pool in the lumen, being directly available for rapid assembly after D1-turnover and PSII re-assembly (Ettinger and Theg, 1991; Hashimoto et al., 1996; Hashimoto et al., 1997). While the involvement of the three proteins in supporting oxygen evolution and in the modulation of inorganic photosynthetic cofactors (manganese, calcium and chloride) has been studied, they have also recently been proposed to have other additional functions (Bricker and Frankel, 2011). The PsbO and PsbP proteins are
required for assembly and stability of the PSII core in normal growth conditions (Ifuku et al., 2005; Yi et al., 2005; Yi et al., 2007), while PsbQ appears to be more specifically required during low light conditions (Yi et al., 2006). Interestingly PsbP was shown to bind manganese (Bondarava et al., 2007), something which may also be highly relevant for members of the PsbP-domain protein family, such as PPD6, which has been characterized in this thesis (Paper V). In order for PSII to function properly, the correct assembly of its constituents is essential. One factor which is required for the assembly is the luminal HCF136 protein. This was revealed by studying *Arabidopsis thaliana* deletion mutants lacking the HCF136 protein, which do not form any PSII reaction centers (Plucken et al., 2002). Several other luminal proteins have also been implicated in the repair and assembly of PSII, including the 18.3 kDa protein and Psb27 (Chen et al., 2006; Sirpio et al., 2007; Liu et al., 2011).

**Electron carriers** - The Arabidopsis lumen contains two isoforms of the Cu²⁺ binding electron carrier plastocyanin, PLAS1 and PLAS2, encoded in the nuclear genome by the *Pete1* and *Pete2* genes (Kieselbach et al., 2000). PLAS2 is the major isoform of the protein, constituting approximately 90 % of the total plastocyanin content in the lumen, while the minor form, PLAS1, accounts for the remaining 10 %. In a thorough analysis of *Pete1* and *Pete2* single and double knock-out mutants, as well as over-expression lines, Pesaresi and colleagues came to the conclusion that the two isoforms are functionally equivalent (Pesaresi et al., 2009). The two proteins have 82 % sequence identity and 92 % sequence similarity, with the most prominent difference being the substitution of tyrosine 33 in PLAS1 to a phenylalanine in PLAS2. In potato this residue was found to be important for interactions with redox partners, but the consequence of the substitution in the Arabidopsis isoforms, if any, is still unclear (Haehnel et al., 1994). A second alternative electron carrier between cyt *b₅f* and PSI in the lumen may be cytochrome *c₆A*. While the silencing by RNA interference of cytochrome *c₆A* alone did not give rise to any detectable phenotype, silenced cytochrome *c₆A* in a double plastocyanin silenced mutant background was lethal, suggesting that cytochrome *c₆A* could replace plastocyanin as electron carrier (Gupta et al., 2002a). This interesting finding was however shortly after challenged by Weigel and co-workers, who could not observe any complementation by cytochrome *c₆A* in their double plastocyanin knock-out mutant (Weigel et al., 2003). This disagreement was further addressed using *in vitro* studies, which showed that cytochrome *c₆A* could not substitute plastocyanin due to a much lower redox midpoint potential (Molina-Heredia et al., 2003). In a review discussing the contradicting results, Howe et al. suggested that cytochrome *c₆A* instead may have a regulatory role in lumen redox signaling, and a hypothesis was proposed where cytochrome *c₆A* catalyzes the formation of disulphide bridges in the lumen space (Howe et al., 2006; Schlarb-Ridley et al., 2006). In a recent study of the folding kinetics and stability of cytochrome *c₆A*, it was however concluded that the proteins
Table 1. Experimentally verified proteins located in the thylakoid lumen of Arabidopsis thaliana.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Gene locus (TAIR)</th>
<th>Proposed function(s)</th>
<th>Import pathway</th>
<th>References</th>
<th>In this thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipocalin domain proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VDE</td>
<td>AT1G08550</td>
<td>Xanthophyll cycle, photoprotection</td>
<td>Sec</td>
<td>A, Niyogi et al. 1998, Arnoux et al. 2009</td>
<td>Paper II</td>
</tr>
<tr>
<td>AtCHL</td>
<td>AT3G47860</td>
<td>Oxidative stress, drought induced</td>
<td>Sec</td>
<td>Levesque-Tremblay et al. 2009</td>
<td></td>
</tr>
<tr>
<td>Photosystem II assembly and repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCF136</td>
<td>AT5G23120</td>
<td>Assembly of PSII</td>
<td>Tat</td>
<td>A, B, Meurer et al. 1998, Plucken et al. 2002</td>
<td>Paper I</td>
</tr>
<tr>
<td>PSb27</td>
<td>AT1G03600</td>
<td>Assembly of PSII</td>
<td>Tat</td>
<td>B, Chen et al. 2006, Liu et al. 2011</td>
<td></td>
</tr>
<tr>
<td>Photosystem II subunits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSbO1</td>
<td>AT5G66570</td>
<td>OEC subunit O</td>
<td>Sec</td>
<td>A, B, Bricker and Frankel 2011 (review)</td>
<td>Paper II</td>
</tr>
<tr>
<td>PSbO2</td>
<td>AT5G50820</td>
<td>OEC subunit O, 2nd isoform</td>
<td>Sec</td>
<td>A, B</td>
<td>Paper II</td>
</tr>
<tr>
<td>PSbP1</td>
<td>AT1G06680</td>
<td>OEC subunit P</td>
<td>Tat</td>
<td>A, B, Bricker and Frankel 2011 (review)</td>
<td>Paper I, Paper II</td>
</tr>
<tr>
<td>PSbP2</td>
<td>AT2G30790</td>
<td>OEC subunit P, 2nd isoform</td>
<td>Tat</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>PSbQ1</td>
<td>AT4G21280</td>
<td>OEC subunit Q</td>
<td>Tat</td>
<td>A, B, Bricker and Frankel 2011 (review)</td>
<td>Paper I</td>
</tr>
<tr>
<td>PSbQ2</td>
<td>AT4G05180</td>
<td>OEC subunit Q, 2nd isoform</td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>Photosystem I subunits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PsaN</td>
<td>AT5G64040</td>
<td></td>
<td>Tat</td>
<td>B</td>
<td>Paper II</td>
</tr>
<tr>
<td>Electron transport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plastocyanin, major</td>
<td>AT5G20340</td>
<td>Photosynthetic electron transport</td>
<td>Sec</td>
<td>A, B, Pesaresi et al. 2009</td>
<td>Paper I</td>
</tr>
<tr>
<td>Plastocyanin, minor</td>
<td>AT5G176100</td>
<td>Photosynthetic electron transport</td>
<td>Sec</td>
<td>A, B, Pesaresi et al. 2009</td>
<td>Paper II</td>
</tr>
<tr>
<td>Cytochrome c6A</td>
<td>AT5G45040</td>
<td>Electron transport, lumen redox signalling</td>
<td>Sec</td>
<td>Gupta et al. 2002a, Weigel et al. 2003</td>
<td></td>
</tr>
<tr>
<td>PsbP domain proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPL1</td>
<td>AT3G55330</td>
<td>PSII repair</td>
<td>Tat</td>
<td>A, B, Ishihara et al. 2007</td>
<td>Paper I</td>
</tr>
<tr>
<td>PPL2</td>
<td>AT2G39470</td>
<td>NDH-complex subunit</td>
<td>Tat</td>
<td>B, Ishihara et al. 2007</td>
<td>Paper I</td>
</tr>
<tr>
<td>PPD1</td>
<td>AT4G15510</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td>Paper I</td>
</tr>
<tr>
<td>PPD2</td>
<td>AT2G28605</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>PPD3</td>
<td>AT7G76450</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>PPD4</td>
<td>AT7G77090</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>PPD5</td>
<td>AT5G11450</td>
<td>Strigolactone biosynthesis</td>
<td>Tat</td>
<td>A, Roose et al. 2011</td>
<td>Paper I</td>
</tr>
<tr>
<td>PPD6</td>
<td>AT3G56650</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td>Paper II, Paper V</td>
</tr>
<tr>
<td>Unnamed</td>
<td>AT5G27390</td>
<td></td>
<td>Tat</td>
<td></td>
<td>Paper I</td>
</tr>
<tr>
<td>PsbQ domain proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PQL1</td>
<td>AT5G44150</td>
<td>NDH-complex subunit</td>
<td>Tat</td>
<td>D, Soursa et al. 2010, Yabuta et al. 2010</td>
<td></td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Protein Family</th>
<th>Gene ID</th>
<th>Description</th>
<th>Localization</th>
<th>Source</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>PQL2</td>
<td>AT3G01440</td>
<td>NDH-complex subunit</td>
<td>Tat</td>
<td>D, Soursa et al. 2010, Yabuta et al. 2010</td>
<td></td>
</tr>
<tr>
<td><strong>Immunophilins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP13</td>
<td>AT5G45680</td>
<td>Rieske complex formation, PPIase activity</td>
<td>Tat</td>
<td>A, Gupta et al. 2002b, Gopalan et al. 2004</td>
<td></td>
</tr>
<tr>
<td>FKBP20-2</td>
<td>AT3G60370</td>
<td></td>
<td>Tat</td>
<td>A, Lima et al. 2006</td>
<td></td>
</tr>
<tr>
<td>FKBP16-1</td>
<td>AT4G26555</td>
<td></td>
<td>Sec</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>17.8 kDa PPIase</td>
<td>AT4G20810</td>
<td></td>
<td>Tat</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>18 kDa PPIase</td>
<td>AT5G13410</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>17.5 kDa PPIase</td>
<td>AT2G43560</td>
<td>Assembly and stabilization of PSII</td>
<td>Sec</td>
<td>A, B, Fu et al. 2007</td>
<td></td>
</tr>
<tr>
<td>Unnamed PPIase</td>
<td>AT3G10060</td>
<td></td>
<td>Tat</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>CYP38</td>
<td>AT3G01480</td>
<td></td>
<td>Sec</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td><strong>Pentapeptide repeat proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL17</td>
<td>AT3G53490</td>
<td></td>
<td>Sec</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>TL15</td>
<td>AT2G44920</td>
<td>Proposed redox-regulated chaperone</td>
<td>Sec</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>TL20.3</td>
<td>AT1G12250</td>
<td>Proposed redox-regulated chaperone</td>
<td>Sec</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td><strong>Proteases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1-processing protease</td>
<td>AT4G17740</td>
<td>Processing of the D1-subunit of PSII</td>
<td>Sec</td>
<td>A, Anbudurai et al. 1994</td>
<td></td>
</tr>
<tr>
<td>D1-proc. protease-like</td>
<td>AT5G46390</td>
<td></td>
<td>Sec</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>DEG1</td>
<td>AT3G27925</td>
<td>ATP independent serine protease</td>
<td>Sec</td>
<td>A, B, Chassin et al. 2002</td>
<td></td>
</tr>
<tr>
<td>DEG5</td>
<td>AT4G18370</td>
<td>ATP independent serine protease</td>
<td>Tat</td>
<td>A, B, Sun et al. 2007</td>
<td></td>
</tr>
<tr>
<td>DEG8</td>
<td>AT5G39830</td>
<td>ATP independent serine protease</td>
<td>Tat</td>
<td>A, Sun et al. 2007</td>
<td></td>
</tr>
<tr>
<td><strong>Other proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrxQ</td>
<td>AT3G26060</td>
<td>Peroxiredoxin</td>
<td>Sec</td>
<td>Pettersson et al. 2006</td>
<td></td>
</tr>
<tr>
<td>C326</td>
<td>AT3G03630</td>
<td>S-sulfocysteine synthase activity</td>
<td>Sec</td>
<td>Bermudez et al. 2012</td>
<td></td>
</tr>
<tr>
<td><strong>No predicted domains or function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL16</td>
<td>AT4G02530</td>
<td></td>
<td>Tat</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>17.9 kDa protein</td>
<td>AT4G24930</td>
<td></td>
<td>Sec</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>15.0 kDa protein</td>
<td>AT5G52970</td>
<td></td>
<td>Sec</td>
<td>A, B</td>
<td></td>
</tr>
<tr>
<td>TL19</td>
<td>AT3G66325</td>
<td></td>
<td>Sec</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Unnamed</td>
<td>AT5G42765</td>
<td></td>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

disulphide bridge instead probably serves a structural, rather than a catalytic role, leaving the question of the proteins function in the lumen still open (Mason et al., 2012).

**Proteases** - Five proteases have been experimentally identified in the chloroplast lumen, two carboxyl-terminal processing peptidases (Ctp) and three Deg family proteases. When the PSII reaction center D1 subunit is synthesized and incorporated into the thylakoid membrane it contains an 8-16 amino acids long C-terminal extension. This extension is removed by a D1 processing protease. This process is necessary for the correct assembly of PSII, as demonstrated in the cyanobacterium *Synechocystis* PCC 6803, where mutants lacking CtpA, a homolog of the lumen Ctps, were unable to perform oxygen evolution (Anbudurai et al., 1994). The three luminal Deg proteins, Deg1, Deg5 and Deg8 are members of the ATP independent serine protease family (Kieselbach and Funk, 2003). Deg1 was found to associate tightly with the thylakoid membrane, and in an *in vitro* analysis using recombinant Deg1 protein, it was shown that Deg1 was able to degrade the soluble lumen proteins plastocyanin and PsbO (Itzhaki et al., 1998; Chassin et al., 2002). In Arabidopsis, Deg5 and Deg8 have been shown to form heterooligomeric hexameric complexes, and recombinant Deg8 protein was also demonstrated to be proteolytically active towards the D1 reaction center protein (Sun et al., 2007).

**Immunophilins** - One of the remarkable features of the lumen proteome is the large group of immunophilins present. Originally defined as being receptors for immunosuppressive drugs, the ubiquitous immunophilin proteins were later shown to exhibit peptidyl-prolyl cis-trans isomerase (PPIase) activity, i.e catalyzing the interconversion between *cis* and *trans* peptide bonds in proline residues. Out of the 17 predicted chloroplast immunophilins in Arabidopsis, all except for one contain a typical bipartite transit peptide, which targets them to the thylakoid lumen (He et al., 2004). While the PPIase activity implies a role for the immunophilins in protein folding and protein trafficking, the precise physiological role of the majority of the luminal members is not yet known. A few of the luminal immunophilins have however been characterized. The luminal immunophilin of highest molecular mass is CYP38, a multidomain protein containing a C-terminal immunophilin domain as well as an N-terminal leucine zipper domain and a central acidic region (He et al., 2004). Its homolog in spinach, TLP40, was the first luminal immunophilin to be characterized. It was co-purified together with a thylakoid bound PP2A-like protein phosphatase which had its phosphatase domain facing the stroma (Fulgosi et al., 1998). The PP2A-like protein was shown to be involved in the de-phosphorylation of PSII reaction center proteins, and its reversible binding to TLP40 was proposed to regulate its phosphatase activity (Vener et al., 1999). Analysis of T-DNA knockout mutants of Arabidopsis CYP38 also later showed that CYP38 is required for the assembly and stabilization of
PSII (Fu et al., 2007). The single domain lumenal immunophilin CYP20-2 is light regulated and its spinach homolog TLP20 has been suggested to be the general folding catalyst in the lumen compartment (Edvardsson et al., 2003; Romano et al., 2004). Additionally the FKBP13 protein was the first lumenal immunophilin shown to be redox regulated (Gopalan et al., 2004). The immunophilin FKBP13 interacts with the Rieske protein, a component of cytochrome $b_6f$ in the photosynthetic electron transport chain (Cramer et al., 1997; Gupta et al., 2002b). The determination of the crystal structure of FKBP13 revealed a pair of disulphide bonds, and reduction of these disulphide bonds by thioredoxin resulted in the loss of PPIase activity (Gopalan et al., 2004). Following this discovery a second lumenal immunophilin, FKBP20-2, which also contained a disulphide bond, was characterized and determined to be required for the proper accumulation of PSII supercomplexes. Interestingly FKBP20-2 however only showed 1/500 of the PPIase activity observed for FKBP13, and the reduction of the disulphide bond in this case did not have any effect on the PPIase activity (Lima et al., 2006). FKBP20-2 lacks three of the five amino acids suggested to be required for PPIase activity, which probably explains its low activity compared to FKBP13. This also raised the question of the predicted PPIase activity of the remaining lumenal PPIases. A comprehensive sequence analysis of all the lumenal immunophilins, concluded that only FKBP13 and CYP20-2 contain all five required amino acid residues, and a subsequent study of Arabidopsis FKBP13/CYP20-2 double knockout mutants showed that these two proteins indeed were responsible for all PPIase activity in the thylakoid lumen (Edvardsson et al., 2007; Ingelsson et al., 2009). Another recent finding was that CYP20-2 also is an auxillary protein of the chloroplast NAD(P)H dehydrogenase complex, a low abundance thylakoid complex which functions in PSI cyclic electron flow and chlororespiration (Sirpio et al., 2009). In conclusion the functions of the still uncharacterized immunophilins in the thylakoid lumen appear to be different to that which has been predicted, and they need to be investigated outside the context of their PPIase activity.

**Protective enzymes** - One of the best characterized proteins in the thylakoid lumen is violaxanthin de-epoxidase (VDE). The protein is a component of the xanthophyll cycle and catalyzes the conversion of the pigment molecule violaxanthin into antheraxanthin and zeaxanthin. Its discovery and characterization have been extensively reviewed (Hieber et al., 2000). It is one of the lumen proteins of highest molecular mass, 43 kDa for the mature signal sequence less form, and it consists of a lipocalin domain and a cysteine rich N-terminal domain. The important role of VDE in plant photosynthesis was shown in an analysis of knock-out mutants in Arabidopsis, where the xanthophyll cycle was suggested to play a central role in the dissipation of excess absorbed light energy (Niyogi et al., 1998). VDE activity is regulated by the lumen pH, and the protein uses ascorbate as a cofactor (Muller-Moule et al., 2002). This suggests the presence of ascorbate within the thylakoid
lumen, which was shown by Foyer and Lelandais, who estimated the average concentration to 3.8 mM in the chloroplast lumen of pea leaf mesophyll cells (Foyer and Lelandais, 1996). Ascorbate is an essential substrate for another type of protective enzyme, ascorbate peroxidases, which scavenge hydrogen peroxide. Located in the lumen is a protein designated as an ascorbate peroxidase and referred to as APX4 (ascorbate peroxidase 4). Recent functional and structural studies have however proposed that it in fact is not an ascorbate peroxidase and that it, although retaining the overall fold of ascorbate peroxidases, does not possess any peroxidase activity and is missing several key amino acids in the active site (Granlund et al., 2009; Lundberg et al., 2011). A thioredoxin dependent peroxiredoxin type peroxidase, PrxQ, was also proposed to be lumen localized, but its precise function still remains unclear (Petersson et al., 2006).

Other proteins - Besides the proteins briefly described above, the lumen also contains representatives of two prominent protein families, three pentapeptide repeat proteins of previously unknown function and nine PsbP-domain proteins. Lumenal members of these protein families have been structurally and functionally characterized in Papers IV and V of this thesis and are discussed in detail in later chapters summarizing that work. The remaining members of the lumen proteome share no sequence similarity to any known proteins from bacteria, animals or plants and still remain uncharacterized. They are an intriguing prospect for future research and the determination of their functions may reveal novel roles for the chloroplast lumen.
Aim of this thesis

While of some of the proteins in the thylakoid lumen have been characterized, there are still many which have no proposed function. These proteins may play significant roles in regulating photosynthesis or other cellular processes and the characterization of their specific functions is therefore of great importance.

General aim

To increase our understanding of the function of the thylakoid lumen. First globally by studying light-regulation of the lumen proteome and by identifying potential targets of the disulphide reductase thioredoxin. Secondly by functional and structural characterization of individual lumen proteins.

Specific aims

Paper I To study the changes in abundance of lumen proteins during the day/night cycle, as well as identify patterns of transcriptional co-expression between genes encoding lumen proteins.

Paper II The identification of putative targets of thioredoxin in the thylakoid lumen, and using this information to gain new insights into functions of luminal proteins.

Paper III Determination of the sub-organelar localization of all known thioredoxin and thioredoxin-like proteins within leaf chloroplasts.

Paper IV Structural and functional characterization of the luminal pentapeptide repeat proteins TL15, TL17 and TL20.3.

Paper V Cloning, expression, purification and obtaining diffraction quality protein crystals of the PsbP-domain protein PPD6.
Results and discussion

Diurnal regulation of the lumen proteome (Paper I)

One way to gain functional information for the lumen proteins is to analyze their regulation in different conditions and developmental stages. In this first study we have examined two basic aspects of the lumen proteome from a global perspective. The first question we asked was how the different lumen proteins change in abundance during a normal day/night cycle, and this question we addressed using difference gel electrophoresis technology (DIGE). The second question was if we could identify any unique patterns of transcriptional co-expression among groups of lumen protein genes, in order to give us an indication of their function.

Diurnal changes in lumen protein expression - Using the original preparation procedures, an isolation of the Arabidopsis thaliana lumen proteome takes approximately six hours (Kieselbach et al., 1998). In order to 'trap' the proteome at a specific state or in a certain condition the extraction procedure must be fast in order to reduce changes in the proteome after the trapped stage. To study the diurnal regulation of the lumen proteome we therefore first needed to modify and simplify our isolation method. By omitting many cumbersome thylakoid membrane washing steps, the isolation time was reduced to two hours from the normal six hours. These wash steps are used to remove as many soluble stroma proteins as possible from the thylakoid membrane fraction, so that a very pure lumen fraction is obtained when the thylakoids are finally ruptured. Not performing these steps consequently leads to more contamination of the lumen fraction by stromal proteins. In other words a significant increase in isolation speed was gained at the expense of purity.

Four independent lumen isolations from 16 h dark-adapted plants and four isolations from 8 h light-adapted plants were first performed. The isolated lumen fractions were labeled with minimal labeling Cy3 and Cy5 fluorescent dyes, according to a dye-swap experimental design (Table 2 in Paper I). An internal standard sample was created by combining equal amounts of all samples in the experiment and subsequently labeling the standard sample with Cy2 dye. Differentially labeled lumen samples were combined and separated first by iso-electric focusing (IEF) and in the second dimension by denaturing polyacrylamide gel electrophoresis (SDS-PAGE). Following gel separation, image acquisition and data normalization using the internal standard, 31 protein spots exhibiting differential abundance between light-adapted and dark-adapted plants were observed. Twenty-nine of the protein spots were identified by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), representing 15 lumenal proteins and seven proteins of stromal or other origin (Table 2, Table 3 in
Interestingly all the identified lumen proteins exhibited an increased abundance in light-adapted plants. Among the identified proteins were the extrinsic components of PSII, PsbP1 and PsbQ2, as well as the electron transport protein plastocyanin (major form). In addition two proteins involved in the assembly and maintenance of PSII; HCF136 and the PsbP-like protein PPL1, were identified (Meurer et al., 1998; Plucken et al., 2002; Ishihara et al., 2007), together with eight lumen proteins of unknown or predicted function. Most of the identified proteins exhibited an increased abundance in light-adapted plants in the range of 1.5-2.0 fold, as compared to the dark-adapted plants. Here it needs to be stressed that fold changes measured in DIGE experiment are only approximations and depend much on the downstream analysis procedure. Depending on the spot border definition parameters used, different image analysis programs will report varying values of fold change. It can however clearly be concluded that there are no individual lumen proteins which show a much higher abundance increase than other lumen proteins. Rather the lumen proteome as a whole appears to generally increase in abundance to the same extent. While the average fold change appears low, only 1.6-fold on average, it is in fact significant considering the crowded lumen environment.

Table 2. Lumen proteins exhibiting increased abundance in light-adapted as compared to dark-adapted Arabidopsis plants

<table>
<thead>
<tr>
<th>Protein description</th>
<th>Gene locus (TAIR)</th>
<th>Nr. of spots</th>
<th>Average ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCF136</td>
<td>At5g23120</td>
<td>2</td>
<td>1.55</td>
</tr>
<tr>
<td>PsbP1</td>
<td>At1g06680</td>
<td>3</td>
<td>1.49</td>
</tr>
<tr>
<td>PsbQ2</td>
<td>At4g05180</td>
<td>3</td>
<td>1.82</td>
</tr>
<tr>
<td>Plastocyanin, major form</td>
<td>At1g20340</td>
<td>2</td>
<td>1.50</td>
</tr>
<tr>
<td>PPD5</td>
<td>At5g11450</td>
<td>1</td>
<td>1.46</td>
</tr>
<tr>
<td>PPD3</td>
<td>At1g76450</td>
<td>1</td>
<td>1.59</td>
</tr>
<tr>
<td>PPL1</td>
<td>At3g55330</td>
<td>1</td>
<td>1.56</td>
</tr>
<tr>
<td>17.5 kDa PPIase</td>
<td>At2g43560</td>
<td>1</td>
<td>1.58</td>
</tr>
<tr>
<td>17.9 kDa unknown function</td>
<td>At4g24930</td>
<td>1</td>
<td>1.63</td>
</tr>
<tr>
<td>Unknown function</td>
<td>At5g42765</td>
<td>1</td>
<td>1.59</td>
</tr>
<tr>
<td>TL17 pentapeptide</td>
<td>At5g53490</td>
<td>1</td>
<td>1.48</td>
</tr>
<tr>
<td>TL15 pentapeptide</td>
<td>At5g44920</td>
<td>1</td>
<td>1.25</td>
</tr>
<tr>
<td>TL16 unknown function</td>
<td>At4g02530</td>
<td>1</td>
<td>1.38</td>
</tr>
<tr>
<td>15 kDa unknown function</td>
<td>At5g52970</td>
<td>1</td>
<td>1.48</td>
</tr>
<tr>
<td>Proposed lumen protein</td>
<td>At5g27390</td>
<td>1</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Uniform transcriptional regulation of lumen protein genes - In order to complement our proteomic work we also examined gene expression profiles of lumen protein genes using publicly available microarray data. A very common way of using this type of data is examining how genes of interest are expressed in different tissues or under certain environmental or developmental conditions. For example, observing that a gene of interest is
upregulated in a certain abiotic stress condition may provide a starting point for functional analysis of that particular gene. While this is certainly a very useful approach, my personal experience working with analysis of experimental microarray data is that it is very noisy, especially for low-to medium expressed genes, and conclusions drawn from individual experiments should be taken with care. This is especially true for publicly available sources, where details regarding the experimental procedures used may be unavailable. A more robust approach is to study co-expression patterns of genes. In other words to instead ask the question: which other genes are regulated in the same way as my gene of interest, across different environmental conditions, developmental stages and tissue types. This is a very powerful method for identification of novel genes involved in different biological processes (Aoki et al., 2007; Ishihara et al., 2007; Yonekura-Sakakibara et al., 2007; Saito et al., 2008). Put simply, expression values for a gene of interest, for example the gene encoding the lumenal TL15 pentapeptide repeat protein, are obtained from hundreds of microarray experiments, representing a wide range of different samples and conditions. The corresponding expression data is obtained from a second gene, for example encoding the TL17 pentapeptide repeat protein, and they are plotted against each other (Figure 7). A correlation coefficient (r-value or CC) is then calculated, reflecting how similar the expression patterns for the two genes are. Two genes which have identical expression profiles across the different microarray experiments would have a correlation (r-value) of 1, while totally unrelated genes would have a value close to 0. Calculating the correlation coefficient for a gene of interest against all genes in the genome

Figure 7. Co-expression analysis. (A) The observed expression level (y-axis) for the genes encoding the lumenal pentapeptide proteins TL17 and TL15 in 1053 microarray slides (x-axis). It is apparent that although TL17 is generally expressed at a higher level than TL15, they follow the same pattern of expression across the microarray experiments, and hence can be considered to be co-expressed genes. (B) Expression of TL17 (y-axis) vs. TL15 (x-axis). The pearson correlation coefficient (r) is equal to the square root of the coefficient of determination obtained upon performing a linear fitting of the data points.
Figure 8. Clustered co-expression correlation heat map of genes encoding thylakoid lumen proteins, as in Paper I. High co-expression is indicated by yellow color, while blue color represents low co-expression.

yields information on which genes, of known and unknown function, that are expressed in the same way. Performing this analysis for a couple of selected lumen protein genes, interestingly returned a list of co-expressed genes which contained many other lumen protein genes. Comprehensively investigating this further, we examined the level of co-expression between all lumen protein genes. This confirmed our hypothesis that almost all lumen protein genes were to a large extent co-expressed, which could be visualized as a clustered correlation heat map (Figure 8). We could also observe physiologically sensible sub-groupings among the lumen protein genes, for example clustering of the extrinsic components of PSII, with each other and with plastocyanin, and clustering of the two pentapeptide repeat family proteins TL15 and TL17. Highly co-expressed with the extrinsic PSII component genes is also the gene encoding a 18.3 kDa protein, which was recently implicated in PSII repair (Sirpio et al., 2007). Using the results from
the co-expression analysis of the lumen protein genes, we identified novel genes which were co-expressed with the known lumen genes. Among these genes there was an interesting representation of genes encoding chloroplast ribosome proteins and redox-signalling proteins, as well as many genes encoding proteins of so far unknown function.

In summary our results showed that at least fifteen lumen proteins exhibited increased expression levels in light-adapted Arabidopsis plants. Among these proteins were several of the most abundant lumen proteins, including PsbP1, PsbQ2, plastocyanin and HCF136. Additionally our analysis of publicly available gene expression data showed that the lumen proteome as a whole may be uniformly regulated at the transcriptional level, implying that most lumen proteins are involved in processes that take place during photosynthesis.
Identification of thioredoxin targets in the chloroplast lumen (Paper II)

While the functions of many chloroplast lumen proteins still remain unknown, a significant discovery contributing to increasing our understanding was the finding by several groups that a number of luminal proteins could interact with thioredoxin in vitro. In a structural study of the immunophilin FKBP13, the authors observed a pair of disulphide bonds which were essential for the proteins peptidyl-prolyl isomerase activity. They further showed that these disulphide bonds could be reduced by both E.coli and chloroplast m-type thioredoxin, which led to a loss of activity (Gopalan et al., 2004). Additional indications of thiol redox pathways in the lumen were provided by proteomic screening studies of thioredoxin targets in Arabidopsis thaliana leaf extracts, which identified the two extrinsic PSII components PsbO1 and PsbO2, as well as the pentapeptide repeat protein TL17, as being putative targets (Marchand et al., 2004; Marchand et al., 2006). While systematic screening studies of other chloroplast compartments, including the envelope membrane, the stroma and the thylakoid membrane, for putative thioredoxin targets had been performed previously (Buchanan and Balmer, 2005; Lindahl and Kieselbach, 2009; Montrichard et al., 2009), no similar studies had focused specifically on the soluble proteome of the chloroplast lumen.

We therefore applied two complementary proteomic approaches to identify novel luminal thioredoxin targets (Figure 9). In the first we utilized thiol specific fluorescence labeling and differential alkylation. The procedure started with isolation of Arabidopsis thaliana lumen extract as described before (Schubert et al., 2002), followed by blocking of free cysteines in the protein extract by alkylation with N-ethylmaleimide (NEM). Upon subsequent incubation of the alkylated lumen fraction with E. coli thioredoxin, disulphide bonds in putative target proteins were reduced, thus exposing new free thiols. The reduced target proteins were then labeled using either a thiol specific fluorescent probe, monobromobimane (MBB), as previously described (Wong et al., 2003; Wong et al., 2004), or a non fluorescent probe such as iodoacetamide (IAM) or dimethylacrylamide (DMA). The complex protein mixture was finally separated by 2D-gel electrophoresis, and the probe was detected either visually in the case of the MBB fluorescent probe, or by mass spectrometry in the case of the non-fluorescent probes. The second approach used recombinant monocysteinic TrxA from Synechocystis sp. PCC6803, where the second cysteine in the active site was changed to a serine by site-directed mutagenesis. This method has in a similar manner been used previously in several studies to identify thioredoxin targets (Motohashi et al., 2001; Balmer et al., 2003; Lindahl and Florencio, 2003; Marchand et al., 2006; Motohashi and Hisabori, 2006). Upon binding to target proteins, the engineered monocysteinic TrxA forms
Figure 9. Schematic overview of the experimental procedures used for identification of thioredoxin targets in isolated thylakoid lumen fractions from Arabidopsis thaliana. (A) Fluorescence labeling and differential alkylation using thiol specific probes. (B) Trx affinity chromatography using poly-histidine tagged monocysteinic TrxA.
stable mixed disulphides with target proteins. A poly-histidine tag on TrxA is then used to bind the mixed disulphide complexes to an immobilized metal affinity column. Following washing to remove unspecific binding, the thioredoxin-target complexes are eluted from an affinity column using imidazole and separated by non-reducing denaturing gel electrophoresis. The lane from the gel is excised, incubated with a reducing agent and separated again by gel electrophoresis, this time in reducing conditions. Off diagonal proteins represent putative TrxA targets and these are subsequently identified by mass spectrometry. An overview of the two procedures outlined above is shown in Figure 9.

**Targets identified by fluorescence labeling, differential alkylation and Trx affinity chromatography** - Using MBB labeling for visual detection of target proteins allowed us to identify ten putative target proteins (Figure 1 in Paper II). These included the previously proposed targets FKBP13, PsbO1, PsbO2 and TL17 as well as novel targets (Table 3). Due to limits in the sensitivity of the MBB probe, and the low relative abundance of many lumen proteins on 2D-gels, we increased the sensitivity of the experiment by using non-fluorescent probes and mass spectrometry for target identification. Although highly sensitive, this procedure has two main drawbacks: first it is tedious, as all visible spots from the 2D-gels need to be carefully analyzed by mass spectrometry; secondly it requires that peptides containing labeled cysteine amino acids formed during tryptic digestion are of suitable size for mass spectrometry and hence can be identified. This however does have the advantage that it is possible to deduce which cysteines in the target protein have been reduced by thioredoxin. Using the differential mass labeling procedure with two different probes we could confirm nine targets identified by fluorescent labeling, identify one additional target protein, FKBP20-2, and gain information regarding which cysteines had been reduced (Figure 2 in Paper II). In order to confirm our results by an independent method, and to identify more potential targets, we applied the immobilized TrxA target trapping procedure. Separation of the eluted TrxA-target protein complexes by 2D non-reducing/reducing gel electrophoresis led to the identification of fifteen luminal targets, out of which eight were not identified in our previous experiments (Figure 3 in Paper II). In total we identified 19 putative thioredoxin targets in the lumen, constituting more than 40% of the lumen proteins typically detected in our lumen isolations (Schubert et al., 2002; Kieselbach and Schroder, 2003). Eight of the targets have been previously detected in studies examining interactions with thioredoxin: the extrinsic PSII proteins PsbO1, PsbO2 (Lemaire et al., 2004; Marchand et al., 2004; Balmer et al., 2006; Marchand et al., 2006) and PsbP1 (Balmer et al., 2006; Marchand et al., 2006), the luminal PSI protein PsaN (Balmer et al., 2006; Motohashi and Hisabori, 2006), the immunophilins FKBP13 and FKBP20-2 (Gopalan et al., 2004; Lima et al., 2006), peroxiredoxin Q (PrxQ) (Motohashi et al., 2001; Marchand et al., 2006) and the luminal pentapeptide repeat protein TL17 (Marchand et al., 2004; Marchand et al., 2006).
The ability of our experimental procedures to also identify these targets strengthens our assumption that the novel targets we have identified do not represent false positives but rather are true novel targets. The novel targets we identified included the xanthophyll cycle enzyme violaxanthin de-epoxidase (VDE), the two D1-processing proteases, the Deg1 and Deg5 proteases, the CYP38 immunophilin, a PsbP-domain protein (PPD6), the two remaining luminal pentapeptide repeat proteins TL15 and TL20.3, the TL29 protein and the TL19 protein. A simplified summary list of all identified luminal targets can be found in Table 3, while a more detailed list is found in Paper II.

Table 3. Thioredoxin targets detected in the thylakoid lumen of Arabidopsis thaliana. The target proteins were identified using the three experimental procedures a) fluorescence labeling, b) differential alkylation and c) Trx affinity chromatography. Targets marked by an asterix (*) were selected for further characterization (Papers IV and V).

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Cysteines in mature protein</th>
<th>Method of identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violaxanthin de-epoxidase</td>
<td>13</td>
<td>a, b, c</td>
</tr>
<tr>
<td>D1 processing protease</td>
<td>5</td>
<td>c</td>
</tr>
<tr>
<td>Putative D1 processing protease</td>
<td>4</td>
<td>c</td>
</tr>
<tr>
<td>Deg1 protease</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>Deg5 protease</td>
<td>2</td>
<td>c</td>
</tr>
<tr>
<td>PsbO1</td>
<td>2</td>
<td>a, b, c</td>
</tr>
<tr>
<td>PsbO2</td>
<td>2</td>
<td>a, b, c</td>
</tr>
<tr>
<td>PsbP1</td>
<td>1</td>
<td>a, c</td>
</tr>
<tr>
<td>PsaN</td>
<td>4</td>
<td>c</td>
</tr>
<tr>
<td>PPD6 (20 kDa PsbP-like protein) *</td>
<td>2</td>
<td>a, b, c</td>
</tr>
<tr>
<td>Cyp38</td>
<td>1</td>
<td>c</td>
</tr>
<tr>
<td>FKB20-2</td>
<td>2</td>
<td>b</td>
</tr>
<tr>
<td>FKB13</td>
<td>4</td>
<td>a, b</td>
</tr>
<tr>
<td>TL17 pentapeptide *</td>
<td>4</td>
<td>a, b</td>
</tr>
<tr>
<td>TL15 pentapeptide *</td>
<td>2</td>
<td>a, b</td>
</tr>
<tr>
<td>TL20.3 pentapeptide *</td>
<td>4</td>
<td>c</td>
</tr>
<tr>
<td>Peroxiredoxin Q</td>
<td>2</td>
<td>a, b, c</td>
</tr>
<tr>
<td>TL29 (Apx4)</td>
<td>2</td>
<td>a, b, c</td>
</tr>
<tr>
<td>TL19</td>
<td>1</td>
<td>c</td>
</tr>
</tbody>
</table>

Redox regulation of the xanthophyll cycle - One of the identified target proteins, violaxanthin de-epoxidase (VDE), is a component of the xanthophyll cycle and catalyzes the conversion of the pigment molecule violaxanthin into antheraxanthin and zeaxanthin (Hieber et al., 2000). It is a 43 kDa lipocalin domain protein containing 13 cysteine residues, most of which are found in a cysteine rich N-terminal domain. In an analysis of VDE knock-out mutants in Arabidopsis the xanthophyll cycle was shown to play a central role in the dissipation of excess absorbed light energy in plants, typically measured as the non-photochemical quenching (NPQ) component.
of chlorophyll fluorescence (Niyogi et al., 1998). Unlike most other lumen proteins, VDE is well characterized and has a known enzymatic activity, for which assays have been previously developed. This made it an interesting subject for further thioredoxin interaction studies. Although it has been shown thirty years ago that VDE activity is inhibited by the reducing agent DTT (Yamamoto and Kamite, 1972), the potential for enzymatic redox-regulation of the enzyme has not so far been explored. To address this we measured the activity of VDE in our thylakoid lumen fractions, in the presence and absence of the E. coli thioredoxin system (Table 4). In line with our hypothesis, more than 80% of the specific activity of VDE was lost in the presence of thioredoxin. Although the first crystal structure of VDE was recently solved, showing how dimerization at acidic pH may facilitate de-epoxidation of the violaxanthin substrate (Arnoux et al., 2009), the structural model only included the central lipocalin domain and unfortunately did not shed any light on the details of its disulphide bridge network and redox regulation, leaving this question still open.

Table 4. Inhibition of VDE activity by E. coli thioredoxin.

<table>
<thead>
<tr>
<th>VDE activity (nmol violaxanthin s⁻¹ mg protein⁻¹)</th>
<th>Lumen isolation 1</th>
<th>Lumen isolation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only lumen fraction</td>
<td>687 ± 103</td>
<td>655 ± 98</td>
</tr>
<tr>
<td>Lumen fraction and E. coli thioredoxin</td>
<td>132 ± 20</td>
<td>17 ± 3</td>
</tr>
</tbody>
</table>

**Redox mediated proteolysis of lumen proteins** - In both the immobilized TrxA affinity target trapping experiments and in an analysis of luminal protein complexes in the presence of the E. coli thioredoxin system (Figure 4 in Paper II), we noticed slight degradation of the PsbO proteins and TL17 pentapeptide repeat protein. This was interesting because previous work has shown that the proteolytic activity in lumen preparations is low under normal non-reducing conditions, and most proteins are not significantly degraded *in vitro* (Kieselbach et al., 1998). To further investigate the degradation process, we incubated lumen fractions together with E. coli thioredoxin for five hours, observing the degradation process by gel electrophoresis and immunoblotting. While most lumen proteins were unaffected and remained stable, we could observe pronounced degradation of the PsbO proteins and TL17 protein in the presence of the thioredoxin system, while no degradation of these proteins was observed in the non-reduced lumen sample (Figure 5 in Paper II). Levels of the identified thioredoxin target FKBP13 on the other hand remained unchanged. These results indicate two things: first that the reduction of disulphide bonds has a more de-stabilizing effect on certain proteins such as PsbO1, PsbO2 and TL17, than on others such as FKBP13, thereby targeting them for degradation, and secondly that reductive activation of proteases may occur.
The first point is consistent with the previous observations that the disulphide bridge in PsbO from spinach is essential for its structural stability, while the reduction of FKBP13 does not induce any major structural changes in the protein (Gopalan et al., 2006; Nikitina et al., 2008). Recombinant TL17 protein also rapidly precipitates upon expression in the reduced dithiol state (Hall, unpublished data). Our soluble lumen fraction contains no known proteases other than the D1-processing proteases and the three Deg proteases, and contamination by stromal proteases has never been observed in any of our preparations (Schubert et al., 2002). Therefore the degradation of PsbO1, PsbO2 and TL17 is probably due to activity of the lumenal Deg proteases. Redox dependent protease activity of recombinant Deg1, using casein as a substrate, has been reported (Stroher and Dietz, 2008) and it is possible that the single conserved cysteine residue of Deg1 could be involved in the formation of oligomers.

In summary we have thoroughly mapped the soluble chloroplast thylakoid lumen for targets of thioredoxin, identifying in total 19 putative targets, which remarkably constitutes more than 40 % of the detected lumen proteome. We have also shown that the important xanthophyll cycle enzyme VDE, identified as a potential target, has its activity efficiently inhibited in the presence of reduced thioredoxin. Furthermore we could show that reduction of disulphide bonds in lumenal proteins by thioredoxin may lead to their structural destabilization and subsequent degradation by a possibly redox activated Deg protease in the lumen. The next step is to confirm these interactions with thioredoxin for individual proteins and examine how redox affects their function in the thylakoid lumen, something which will be addressed for the pentapeptide repeat proteins and the PsbP-domain protein PPD6 in coming sections.
Localization studies of chloroplastic thioredoxins (Paper III)

In *Arabidopsis thaliana* there are at least 45 genes encoding thioredoxin and thioredoxin-like proteins. 19 of these proteins are predicted, based on sequence analysis, to be localized in plastids (Meyer et al., 2005). The predicted plastid location has also been supported by experimental evidence for six of the thioredoxins, using GFP-fusion constructs and immunocytolocalisation (Eymery and Rey, 1999; Collin et al., 2003; Collin et al., 2004). The precise sub-chloroplastic localization for most of the chloroplast thioredoxins however remains undetermined.

As a first step in increasing our understanding of the complex network of multiple chloroplast thioredoxins, their overlapping target specificities and their interactions with target proteins, we attempted to systematically map the location within the chloroplast for as many of the thioredoxin proteins as possible. For this we used publicly available cDNA clones and *in vitro* chloroplast import assays.

The experimental procedure starts with the synthesis of a radiolabeled form of the protein of interest using a wheat germ cell-free expression system in the presence of a 35S-methionine/35S-cysteine mixture. The radiolabelled precursor proteins are then incubated together with intact chloroplasts, isolated from leaf tips of pea seedlings. During the 20 min incubation in the presence of light and ATP, any chloroplast targeted proteins will enter the chloroplast via the TOC/TIC pathway, have their chloroplast transit peptide removed and be further targeted to their sub-chloroplastic localization. To remove un-imported precursor protein, the intact chloroplasts are incubated for 40 min with thermolysin protease, before being fractionated into stromal and thylakoid sub-compartments. The thylakoid membrane fraction is further incubated with the trypsin protease to test for targeting of proteins to the lumen compartment. All fractions from the experiment can finally be assayed for the protein of interest by gel electrophoretic separation, followed by detection of the radiolabelled protein. As a methodological note, no non-chloroplast proteins have to our knowledge ever been found to import into isolated chloroplasts, suggesting that results from this type of assay provide very clear evidence of chloroplast localization.

*All f-, m-, x- and y-type thioredoxins are targeted to the stroma* - While plastid localization of the classic thioredoxins Trx-f1, Trx-m2, Trx-m3, Trx-x and Trx-y2 has been shown previously using GFP fusion (Collin et al., 2003; Collin et al., 2004), the localization of the other members of these types, namely Trx-f2, Trx-m1, Trx-m4 and Trx-y1 have not been confirmed. Our import assays of all the members of the f-type and m-type thioredoxins, both confirmed the previous results, as well as confirmed that the other members are also localized to plastids, specifically to the stroma compartment (Figure 10 A).
Figure 10. Autoradiograms showing some of the results from chloroplast import assays of predicted chloroplast thioredoxins. Lanes are denoted as: Tr (translated precursor), C (intact chloroplasts), C+ (protease treated chloroplasts), S (stroma), T (thylakoids) and T+ (protease treated thylakoids). (A) Import of (A) the classical Trx-m, Trx-f and Trx-y type thioredoxins, (B) the multidomain APR and CDSP32 proteins and (C) the WCRKC and lilium thioredoxin-like proteins into the stroma (lane S) of isolated pea chloroplasts. (D) Import of HCF164 to the thylakoid membrane (lane T) and its degradation by trypsin treatment (lane T+). Adapted from Paper I.
The APR and CDSP32 thioredoxins are also localized in the stroma - In contrast to the classical thioredoxins, which all have a WCGPC active site motif, there are at least twelve additional genes encoding proteins with thioredoxin domains which have atypical active sites, and are predicted to be plastid targeted. Four of them are larger proteins, where the thioredoxin fold constitutes only a part of a multidomain protein. The three APR proteins contain an N-terminal adenylylsulphate reductase domain followed by a C-terminal thioredoxin domain with a WCPFC motif in the active site. Interestingly, despite the atypical active site, the APR proteins could complement *E. coli* lacking the TrxA gene (Setya et al., 1996). Setya and co-workers also predicted that the APR1 and APR3 proteins were targeted to plastids, but that APR2, which did not have a clear plastid transit peptide, in contrast probably was targeted differently. In our import assays we confirmed that APR1 indeed is targeted to the chloroplast, more specifically to the stroma compartment (and most probably also APR3 based on its high sequence similarity to APR1) and that this is also the case for APR2, thus placing all three proteins in the chloroplast stroma (Figure 10 B). The 32 kDa thioredoxin protein CDSP32 is involved in oxidative stress response during drought conditions and has been well characterized, both functionally and with regard to its location in the plastid stroma (Rey et al., 1998; Eymery and Rey, 1999; Broin et al., 2000; Broin and Rey, 2003; Rey et al., 2005; Tarrago et al., 2010). We confirmed the stromal localization also for this protein using import into isolated chloroplasts (Figure 10 B).

The putative WCRKC and Lillium thioredoxins are also stroma located - Seven genes encoding putative thioredoxins, two WCRKC motif thioredoxins and five so called lillium proteins, with at the time undetermined redox activity and unconventional active site motifs, were also predicted to be plastid localized. We observed that both WCRKC proteins were imported into the chloroplast stroma, as was lillum 5, a representative member of the lillium family (Figure 10 C). Our results were shortly after confirmed in an independent study in which they proved chloroplast localization for all five lillium proteins using GFP fusion constructs (Dangoor et al., 2009). In contrast to us they observed a partitioning of lillium5 between the stroma and thylakoid membrane, a phenomenon which we could not see in the data from our import experiments.

Does the thioredoxin domain of HCF164 face the lumen? - The thioredoxin-like protein HCF164 is the only chloroplast thioredoxin so far which is known to be located in the thylakoid membrane. The HCF164 precursor protein has a luminal bipartite transit peptide, directing it to the thylakoid via the Tat-pathway. Unlike other lumen proteins however, the transit peptide is not cleaved when the protein enters the lumen but rather functions as a membrane anchor for the lumens exposed thioredoxin domain (Lennartz et al., 2001; Motohashi and Hisabori, 2006). In the work by Lennartz and co-workers, later confirmed by Motohasi and Hisabori, the
premise that the thioredoxin domain faces the lumen was based on protease protection assays using thermolysin on intact Arabidopsis chloroplasts. Interestingly in our import assays using pea chloroplasts it appeared that the soluble domain of HCF164 was not protected from protease treatment and that the protein maybe had an alternative topology (Figure 10 D, lane T+). The main difference between our and the other studies is that we used trypsin protease rather than thermolysin in the protection assay. However the data from both the previous studies and our own study appear robust and an explanation for the discrepancy is not obvious.

Although the main goal of this study was to map the sub-chloroplastic location of as many thioredoxin proteins as possible, our secondary goal was to identify a soluble luminal thioredoxin. However, our quite conclusive mapping showed that all soluble chloroplast thioredoxins are located exclusively to the chloroplast stroma. While this implies that the major source of disulphide reductase activity in the lumen may be accounted for by membrane bound proteins such as HCF164, it however does not rule out the possibility that there are as of yet undiscovered soluble thioredoxin-like proteins located in the lumen.
Characterization of the pentapeptide repeat proteins (Paper IV)

The Arabidopsis genome encodes four proteins containing a so called pentapeptide repeat domain. Out of these four, three are localized in the thylakoid lumen: TL15, TL17 and TL20.3 (Kieselbach et al., 1998; Schubert et al., 2002; Friso et al., 2004). Besides their lumen localization an additional feature of these three proteins is that they were also all identified as potential targets of thioredoxin, with their conserved disulphide bridges being reduced in vitro (Paper II). While plants typically have only a few genes encoding pentapeptide repeat proteins in their genome, cyanobacteria have many more (Chandler et al., 2003). For the cyanobacterium *Cyanothece* sp. PCC 51142 for example, as many as 35 pentapeptide repeat proteins genes have been identified, encoding proteins localized in the lumen/periplasm, cytosol and plasma membrane, (Buchko et al., 2006). Following the identification of the first pentapeptide repeat domain protein by Black et al. in 1995, the pentapeptide repeat family of proteins was defined as containing at least eight tandem repeats of five amino acids of the approximate consensus sequence A(D/N)LXX, and a structural model constituting a three sided beta-helix was proposed by Bateman et al. in 1998 (Black et al., 1995; Bateman et al., 1998). While an interesting secondary role for the pentapeptide repeat protein RfrA from *Synechocystis* PCC 6803 in manganese uptake was shown (Chandler et al., 2003), it was not until the first structure determination of a pentapeptide repeat protein, MfpA from *M. tuberculosis*, that any function could be assigned to the pentapeptide domain (Hegde et al., 2005). Upon solving the structure of MfpA, the authors observed that the pentapeptide repeats indeed formed a beta-helix structure as predicted by Bateman et al. (1998), although it was four- and not three sided. Interestingly Hegde et al. (2005) also showed that MfpA could bind to DNA gyrase by way of DNA mimicking, thereby conferring resistance against the quinolone class of antibiotics. Since then a number of other pentapeptide repeat proteins with proposed DNA mimicking function have been further identified in pathogenic bacteria (Hegde et al., 2011; Vetting et al., 2011a; Vetting et al., 2011b). For the cyanobacterial and plant pentapeptide repeat proteins, their functions however still remain unknown. It is unlikely that many have DNA mimicking function due to the variation in the predicted subcellular locations, with many being localized in compartments devoid of nucleic acids, such as the thylakoid lumen. We therefore characterized the luminal pentapeptide repeat proteins of Arabidopsis and examined the regulatory role their disulphide bridges may play in their proposed function.

**Common domain structure** - The luminal pentapeptide repeat proteins share relatively low sequence identity, 27 % for TL17:TL15, 28 % for TL15:TL20.3 and 16 % for TL17:TL20.3. They all however are predicted to contain a pentapeptide domain consisting of approximately 20 tandem repeats (Figure 11). They also all have two conserved cysteine residues in the C-terminal part of the protein which probably form regulatory disulphide...
bonds (Paper II). Compared to TL15, TL17 has an extended N-terminal region, containing an additional two disulphide forming cysteine residues, and TL20.3 instead has an extended C-terminal region containing two additional cysteines. Besides the common domain structure, the TL15 and TL17 lumenal pentapeptides have also been shown to exhibit diurnal regulation and to generally be co-expressed at the transcriptional level (Paper I). This provides indications that they may be involved in the same processes in the thyalkoid lumen.

High-resolution crystal structure of TL15 - As a first characterization of the luminal pentapeptide repeat proteins, the structure of TL15 was solved by x-ray crystallography. The mature form of the TL15 protein was cloned and expressed in a disulphide promoting E. coli host. Recombinant protein could then be purified to homogeneity using a single step of immobilized metal affinity chromatography (IMAC). Following cleavage of the poly-histidine tag used for IMAC, the protein was crystallized in 200 nL drops containing 0.2 M ammonium sulphate and 30 (w/v) % poly ethylene glycol 8000. Two small crystals were obtained and x-ray diffraction data were collected to a maximum resolution of 1.23 Å. Using the known structure of Rfr32, from Cyanothece sp. PCC 51142 (Buchko et al., 2006), as a search model for initial phase estimation by the molecular replacement method, a structural model of the TL15 protein could be constructed, using a combination of automated and manual methods. TL15 is comprised of a pentapeptide repeat domain, which forms a right-handed quadrilateral beta-helix, and a C-terminal alpha helix.
helix domain containing two alpha helices (Figure 12 A, B). Four consecutive pentapeptide repeats constitute a turn in the beta helix, with the beta helix consisting in total of five turns. The helices in the C-terminal alpha helix domain are connected by a disulphide bond, formed between the two conserved cysteine amino acid residues. The whole C-terminal domain constitutes a cap, covering the C-terminal end of the beta helix. At first glance it appears that reduction of the disulphide bond would 'release' the helices from each other and possibly cause dramatic conformational changes to the C-terminal region. A closer analysis of the loop region connecting the two alpha helices however showed that the loop is connected to the face of the pentapeptide domain by an intricate hydrogen bond network. So while reduction of the disulphide bond surely leads to structural changes, the integrity of the C-terminal cap would probably remain uncompromised. The interior of the pentapeptide beta helix consists mainly of well ordered

Figure 12. Overall topology and structural features of TL15 from A. thaliana. (A) Top-down view, highlighting the C-terminal alpha-helical domain, consisting of two α-helices connected by a disulphide bond between residues C122 and C142. (B) Side-view of the structure. (C) Cut-away view illustrating the well ordered aromatic and aliphatic stacks in the beta-helix interior. (D) Electrostatic surface potential rendering of face 4 (left) and face 2 (right), colored from -0.5 V (red) to 0.5 V (blue)
aliphatic or aromatic stacks, with the central amino acid of each pentapeptide typically being a leucine (50 %) or phenylalanine (35 %) (Figure 12 C). While the structural analysis of the *Nostoc punctiforme* Np275/276 fusion protein identified a continuous internal cavity along the beta helix axis (Vetting et al., 2007), no such cavity can be observed in the TL15 interior. An analysis of atomic displacement parameters, also referred to as temperature factors, showed that the pentapeptide repeat domain is a rigid structure with relatively low temperature factors, while the C-terminal domain is more dynamic, exhibiting lateral thermal motion at an approximate 60° angle from the helical axis of the first alpha helix.

**Chaperone function of the lumenal pentapeptides** - While the x-ray crystallographic analysis of TL15 revealed the fine details of its molecular structure, it did not give any clues to its function in the thylakoid lumen. Based on co-expression analysis and our own hypothesis regarding the general function of lumen proteins we tested the pentapeptide proteins for a possible role in protein folding/assembly. To test for chaperone-like activity, the ability of the TL15, TL17 and TL20.3 proteins to refold a chemically denatured model substrate, MalS, from *E. coli* was assessed as described previously (Malet et al., 2012). Under the *in vitro* conditions used, TL15 and TL20.3 convincingly assisted in the refolding of MalS (Figure 4 in Paper IV). TL17 however showed activity similar to hen egg white lysozyme, suggesting that the activity is only due to unspecific interactions. Upon reduction of disulphide bonds in the pentapeptide proteins using the reducing agent DTT, followed by alkylation with iodoacetamide, we interestingly observed a 1.4 and 3.2-fold increased chaperone activity for TL15 and TL20.3 respectively (Figure 4 in Paper IV). For TL15 this was additionally confirmed using site-directed mutagenesis of one cysteine residue, permanently disrupting the disulphide bridge, which resulted in a 2.4-fold higher activity. Complementary experiments testing if the lumenal pentapeptide proteins could protect a number of model proteins from thermal aggregation however showed that their presence had no effect (data not shown). This implies that TL15 and TL20.3 function as folding chaperones in the lumen, assisting in the correct folding and assembly of other thylakoid proteins, rather than functioning in stress protection.

To summarize we have determined the x-ray crystal structure of the oxidized disulphide form of the pentapeptide repeat protein TL15 from *Arabidopsis thaliana* at 1.3 Å resolution, the highest resolution structure of any pentapeptide repeat protein so far. We further propose that TL15 and TL20.3 may function as redox regulated molecular chaperones, assisting in the folding and proper assembly of proteins in the lumen and thylakoid membrane.
Purification and crystallization of PPD6, a PsbP-domain protein (Paper V)

The thylakoid lumen of Arabidopsis contains at least ten PsbP family proteins, two true PsbP proteins and eight PsbP homologues. Studied using RNA interference lines, true PsbP has been characterized as being required for the proper assembly and function of PSII (Ifuku et al., 2005; Yi et al., 2007; Ido et al., 2009). Also PsbP, but not PsbQ, is essential for normal thylakoid architecture in Arabidopsis (Yi et al., 2009). The eight PsbP homologs can be divided into two groups based on their sequence similarity to the true PsbP proteins, with one group containing two PsbP-like proteins (PPL1 and PPL2), and the other containing six PsbP-domain proteins (PPD1-PPD6). In contrast to PsbP, an analysis of TDNA knockout mutants of PPL1 and PPL2 showed that they were not directly involved in oxygen evolution, but rather had other distinct functions. PPL1 is required for efficient repair of photodamaged PSII while PPL2 is a component of the chloroplast NAD(P)H dehydrogenase complex (Ishihara et al., 2007). Interestingly a recent study of the PsbP domain protein 5 (PPD5) indicated yet another distinct function for a PsbP family protein. While analyzing knockout lines of PPD5 revealed no significant alterations in photosynthetic electron reactions, the mutants exhibited significant developmental and morphological defects (Roose et al., 2011). The plants had both increased lateral root branching and defects associated with axillary bud formation, a so far unique phenotype for a thylakoid lumen protein. The remaining members of the PsbP family in plants still have no proposed functions.

Expression, purification and crystallization of PPD6 - The mature PPD6 protein contains two conserved cysteines, separated by three amino acids, which form a disulphide bridge, and was identified in Paper II as being a potential target of thioredoxin. This is a unique feature of PPD6, as no other PsbP family member in Arabidopsis contains more than a single cysteine in the mature protein sequence. In order to understand the function of PPD6 and also its redox regulation, we wished to determine the structure of the protein by x-ray crystallography. The first step in the structure solution was to establish and optimize expression, purification and crystallization protocols. Signal sequence less PPD6 protein was expressed in a disulphide promoting strain of *E. coli* and in the first step purified by immobilized metal affinity chromatography (IMAC). Subsequent purification using size-exclusion chromatography (SEC) yielded a homogenous protein solution containing monomeric PPD6 protein. Following initial crystallization trials using PPD6 protein containing a poly-histidine tag, an optimized crystallization screen resulted in the formation of star shaped crystals (Figure 13 A). These crystals however could not be used for structure determination due to extensive smearing of the diffraction spots. Cleavage of the poly-histidine tag prior to crystallization instead resulted in the formation of triangular and rhomboidal crystals up to 0.5 mm in length,
grown in 0.1 M Hepes-KOH pH 6.5, 20 % (w/v) PEG 4000 and 0.2 M zinc acetate, using the hanging drop vapor diffusion method (Figure 13 B). These crystals were of good diffraction quality and multiple sets of data were collected at the MAX-lab synchrotron radiation source in Lund, Sweden. An initial analysis of the diffraction data showed that the crystals diffracted to a maximum resolution of 2.1 Å and belonged to space group P2₁, with unit cell parameters $a = 47.0$, $b = 64.3$, $c = 62.1$ Å and $\beta = 94.2^\circ$. Details of data collection and processing statistics can be found in Table 1 of Paper V. Mathews coefficient calculation suggests that PPD6 most probably occurs as a dimer in the crystal, and this has later been confirmed during the structure solution procedure (Hall, unpublished results). An interesting component of the crystallization solution used was 0.2 M zinc acetate. The zinc was required for crystal growth and could not be substituted for other divalent metal ions such as copper or magnesium, which implies specific binding of zinc ions to PPD6. The previously solved structures of cyanobacterial PsbP, as well as PsbP from spinach, also contain bound zinc atoms, which suggests that the residues involved in zinc binding may be conserved throughout the PsbP family (Michoux et al., 2010). The binding of zinc to PPD6 could be used for phase determination by the single wavelength anomalous dispersion (SAD) method, and structure determination of the protein is currently underway using this experimental phase information.

Figure 13. Protein crystals of PPD6. (A) Star-shaped crystals obtained using poly-histidine tagged protein and (B) the triangular/rhomboidal crystals of untagged PPD6 grown in 0.1 M Hepes pH 6.5, 20 % PEG 4000 and 0.2 M zinc acetate used for structure determination.
Summary and conclusions

The proteome of the thylakoid lumen is unique in the sense that it contains no known metabolic enzymes or signaling proteins. Rather the lumen is comprised mainly of only a few different protein families, of which many members have no proposed function. With the work performed in this thesis, we have aimed to increase our understanding of the proteins found in the thylakoid lumen, using global proteomic approaches, as well as studies of individual proteins.

In our analysis of diurnal expression patterns in the lumen proteome, we identified several lumen proteins exhibiting elevated abundance levels in light-adapted Arabidopsis plants. Among the proteins were several with previously implicated roles in the light driven photosynthetic reactions, including the PsbP and PsbQ subunits of the OEC, plastocyanin and the HCF136 and PPL1 proteins (Plucken et al., 2002; Ishihara et al., 2007; Bricker and Frankel, 2011). The increased abundance of these proteins in the light is consistent with their proposed functions and implies that also the increased abundance of those lumen proteins we identified which have so far unknown function is physiologically significant. While the estimated increases in protein levels in our DIGE analysis were ‘only’ in the range of 40-80 %, they are consistent with the model by Kirchoff et al. (2011), where they propose that the distance between opposing thylakoid membranes is approximately doubled in light-adapted plants (Figure 3 in chapter 2). An even greater abundance increase, for example in the range of 5-10 fold would be difficult to envision within the observed expansion of the lumen space. Especially since our analysis identified plastocyanin, PsbP and PsbQ, three proteins of comparatively high abundance, which together constitute the major part of the total protein content in the lumen.

While the identification of proteins showing increased abundance in the light was an expected result, our analysis of transcriptional profiles was a bit more remarkable, showing that almost all genes encoding lumen proteins were regulated in a similar manner. This implies that many of the lumen proteins are in some way involved in the same general cellular process, related to photosynthesis, and gave insight into possible interactions and functional relationships between lumen proteins. It does however not provide evidence regarding relationships with other chloroplast proteins or precise functions. As a strong proponent of the co-expression analysis approach, I believe that additional important insights may be given by expanding the analysis to include all genes, encoding both empirically determined and predicted chloroplast proteins, and to study the co-expression patterns of the lumen protein genes in this context. This type of analysis can be performed for individual genes using online tools, for example ATTED-II (www.atted.jp), Arabidopsis Coexpression Data Mining Tool (www.arabidopsis.leeds.ac.uk),
and recently also Genevestigator (www.genevestigator.com), and this approach has been used to elucidate novel functions for lumen proteins, such as PPL1 and PPL2 (Ishihara et al., 2007). Some of these tools, notably ATTED-II, also have powerful options for analyzing multiple genes simultaneously. The results of such analysis are however often represented as gene networks, which are visualized by a complex pattern of cross connected circles, with each circle representing a gene. While probably one of the most robust methods of co-expression analysis, it becomes difficult to interpret, when very many genes (hundreds to thousands) are analyzed simultaneously. According to my personal opinion, clustered heat map

![Clustered correlation heat map representation of co-expression networks in Arabidopsis chloroplasts. NASCarray data for all genes encoding predicted and known chloroplast proteins according to the TAIR database were used to calculate co-expression correlation coefficients between genes. A high co-expression is represented by red color and a low co-expression by blue. Examples of prominent clusters of co-expressed genes are enclosed by boxes, including a photosystem I cluster, a NDH complex cluster, and a cluster comprised mainly of genes encoding components of the plastid ribosome (enlarged).](image)

**Figure 14.** Clustered correlation heat map representation of co-expression networks in Arabidopsis chloroplasts. NASCarray data for all genes encoding predicted and known chloroplast proteins according to the TAIR database were used to calculate co-expression correlation coefficients between genes. A high co-expression is represented by red color and a low co-expression by blue. Examples of prominent clusters of co-expressed genes are enclosed by boxes, including a photosystem I cluster, a NDH complex cluster, and a cluster comprised mainly of genes encoding components of the plastid ribosome (enlarged).
representations, such as the one used in Paper I, are much more intuitive and easy to interpret, especially for scientists unfamiliar with global gene expression analysis. Large clusters of co-expressing genes can easily be identified and also their relationships to other clusters are directly observable. An example of performing such an analysis on transcriptome data from ~3000 microarrays, publicly available in the NASCArrays database, for all genes encoding chloroplast proteins according to Gene Ontology classification in The Arabidopsis Information Resource (TAIR), encompassing 2628 genes in total, is presented in Figure 14. While a detailed analysis is beyond the scope of this thesis, a few immediate observations can be made. Distinct clusters of highly co-expressed genes are clearly visible, with a typical example being a highly co-expressed Photosystem I cluster, containing many Photosystem I subunits. Another interesting example is a highly co-expressed gene cluster representing many components of the chloroplast NAD(P)H dehydrogenase (NDH) complex. Included in this cluster are the genes encoding PPL2 and PQL1 which have previously been shown to represent luminal NDH subunits (Ishihara et al., 2007; Suorsa et al., 2010; Yabuta et al., 2010). Interestingly additional genes in the NDH cluster include the TL20.3 pentapeptide repeat protein characterized in Paper IV, the luminal Deg1 protease and the HCF164 membrane bound thioredoxin studied in Paper III and discussed more below. A third gene cluster exemplifying possible novel functions for lumen proteins, consists to >60 % of ribosomal protein subunit genes and genes encoding other protein components of the chloroplast translational machinery. Clustered together with these genes are interestingly the gene encoding FKBP20-2, which has been implicated in assembly of PSII supercomplexes (Lima et al., 2006), the gene encoding the D1-processing protease, and also the gene encoding luminal Peroxiredoxin Q (PrxQ). PrxQ is a thioredoxin dependent peroxidase which possibly is involved in oxidative stress protection in the lumen (Petersson et al., 2006). The co-expression analysis however strongly suggests that PrxQ is somehow involved in the chloroplast translation process. Besides the examples I have given above, a more detailed analysis may provide further clues and new insights into the functions of both lumen proteins and the multitude of thioredoxin family proteins studied in Paper III.

Our mapping of thioredoxin targets in the thylakoid lumen revealed 19 proteins, representing all the major protein families in the lumen, including the immunophilins, the PsbP-domain proteins and the pentapeptide repeat proteins. In total more than 40 % of the known lumen proteins in Arabidopsis were identified as potential targets. According to the classical model of thioredoxin mediated regulation, biosynthetic enzymes in the stroma are activated in the light by disulphide reduction via the ferredoxin-thioredoxin system and inactivated in the dark by re-oxidation (Buchanan and Balmer, 2005). In contrast, catabolic enzymes are activated by oxidation in the dark and inactivated by thioredoxin in the light. The signals
controlling the thiol redox state of luminal proteins are however still unclear. The overall paradigm appears to be opposite to that in the stroma, with activation in the light requiring oxidation of cysteines into disulphides rather than reduction. This is exemplified by FKBP13 and violaxanthin de-epoxidase (Gopalan et al., 2004)(Paper II) which are expected to perform their function in the light, and require intact disulphide bridges for activity. Also, the oxygen evolving complex subunits PsbO1 and PsbO2 appear to have disulphide bridges which are essential for their structural stability (Nikitina et al., 2008)(Paper II). In the first proposed model of luminal redox regulation by Buchanan and Luan (2005), they suggest that FKBP13, and perhaps other light activated lumene proteins, may pass through the thylakoid membrane in the reduced state, reduced by the light driven ferredoxin-thioredoxin system in the stroma, and become oxidized upon entering the lumen, either directly via oxygen produced at PSII or via an intermediary redox protein. This would however be a uni-directional process in which oxidized lumen proteins would remain oxidized also in the dark, unless novel redox proteins were discovered. Lending support to the hypothesis that reversible thiol-disulphide exchange reactions may also take within the lumen space, Gopalan et al. (2006) observed that the oxidized and reduced forms of FKBP13 did not differ much structurally, a characteristic typical of redox regulated proteins. Similarly, a thermal unfolding analysis of PPD6 performed by CD spectroscopy, revealed that the oxidized and reduced forms of also this lumen protein, are similar in regards to structural stability (von Sydow and Hall, unpublished) In addition the luminal peroxiredoxin Q is activated by the reduction of its disulphide bond and the luminal pentapeptide repeat proteins TL15 and TL20.3 exhibit higher activity in their reduced forms (Aden et al., 2011)(Paper IV). Since the initial model was presented, several novel redox proteins have indeed been identified, including the thioredoxin domain protein HCF164 and CcdA, which represent components of a DsbD-like system in the thylakoid membrane (Motohashi and Hisabori, 2006, 2010). In the model proposed by Motohashi and Hisabori, HCF164 can transfer reducing equivalents to luminal target proteins from FTR-dependent stromal thioredoxin-m, mediated by CcdA. This model implies that the reduction of lumen target proteins via HFC164 would take place in the light. The lumen is however acidified in the light, with recent measurements suggesting a pH of 5.4-6.0 (Tikhonov et al., 2008), conditions which are unfavorable for thiol-disulphide exchange reactions. At such low pH, below the pKa of typical thiol groups, the non-reactive protonated thiol form (-SH) is favored over the reactive unprotonated thiolate (-S-) form. The thiolate form is required for thioredoxin catalyzed disulphide reduction, implying that the luminal thioredoxin domain of HCF164 would not be catalytically active in the light. Nor so would any other putative disulphide reductase in the lumen be, suggesting that the reduction of lumen proteins may be limited to the dark periods, when the pH is neutral. Another so far unexplored alternative is that the CcdA/HCF164 system may instead be catalyzed in the dark by NTRC, a
recently characterized stromal thioredoxin/thioredoxin reductase, which functions independently of light (Perez-Ruiz et al., 2006; Spinola et al., 2008). Such a model would facilitate the activation of enzymes such as FKBP13 and violaxanthin de-epoxidase in the light, with subsequent inactivation occurring with the onset of darkness. While this is attractive as a model for general disulphide reduction in the lumen, the midpoint redox potential of HCF164 was determined to be -224 mV (Motohashi and Hisabori, 2006), significantly higher than for most stromal thioredoxins, which have midpoint redox potentials in the range -368 mV (Trx-m2) to -336 mV (Trx-x) (Collin et al., 2003). The Populus homolog of the luminal peroxiredoxin PrxQ, which requires thioredoxin mediated reductive activation, has a midpoint redox potential of -325 mV (Rouhier et al., 2003), and as an example could hence not be reduced by HCF164. This suggests that HCF164maybe does not function as a general reductant in the lumen, but rather is limited to its role in the biosynthesis of the cytochrome b_{6f} complex (Lennartz et al., 2001), or other more specific processes. Regarding the re-oxidation of lumen proteins, Cytochrome c_{6A} was initially proposed as a catalyst (Schlarb-Ridley et al., 2006), but this recently was shown to be unlikely (Mason et al., 2012). Two recent studies however identified the gene product of At4g35760 as a member of a novel type of disulphide bond forming proteins, first characterized in cyanobacteria (Goodstadt and Ponting, 2004), localized integrally in the thylakoid membrane (Feng et al., 2011; Karamoko et al., 2011). The protein was named LTO1 (Lumen Thiol Oxidoreductase 1) by Karamoko et al. and it was shown by both groups to have its catalytic site facing the lumen side of the thylakoid membrane. LTO1 knockout mutants showed limitations in electron transfer from PSII and recombinant LTO1 could introduce disulphide bonds in reduced PsbO2 in vitro (Karamoko et al., 2011). The midpoint redox potential of the soluble domain of LTO1 was determined to -180 mV (Furt et al., 2010), further supporting the possibility that it can catalyze the formation of disulphide bonds in luminal proteins in vivo. While several models and proposed components of a regulatory thiol/disulphide exchange network in the lumen have been discussed above, no clear picture, taking into account the differing conditions in the light and in the dark, is as of yet available, and there are still numerous gaps which need to be filled.

One interesting prospect is another type of redox proteins recently discovered in the thylakoid membrane, zinc finger disulphide isomerases. The LQY1 protein contains a transmembrane region anchoring it to the thylakoid membrane and a DnaJ-domain (Lu et al., 2011). The chaperone DnaJ from E. coli has been shown to catalyze disulphide reduction, disulphide formation and disulphide isomerization, in addition to its known function as a chaperone (Decrouychanel et al., 1995). In contrast to DnaJ, LQY1 only contains the cysteine rich zinc finger domain, and not the other domains (J-domain, gly-rich motif and C-terminal domain) found in DnaJ. However, LQY1 was still shown to be able to both break and reform
disulphide bonds in substrate proteins, and was furthermore implicated in the repair process of PSII (Lu et al., 2011). The thiol-disulphide oxidoreductase activity of LQY1 was only demonstrated for unphysiological RNase substrates and not for any PSII subunits or luminal proteins, and it should also be noted that, while the authors in their model proposed that the catalytic zinc finger domain faces the thylakoid lumen, this was only based on a topology prediction and not on experimental evidence.

Interestingly, a gene encoding a second DnaJ-domain zinc finger protein (At2g34860) was found to co-express with the luminal pentapeptide genes and the LQY1 gene (Paper IV). Upon manually examining the chloroplast targeting peptide for this protein and its orthologs from several other plant species, a conserved twin arginine motif can be identified (Figure 15). The twin arginine motif is followed by a hydrophobic core region and three residues of the consensus sequence A-X-A, which are attributes typical for signal peptides targeting proteins to the lumen via the Tat pathway (Albiniak et al., 2012). A luminal location for this protein was also proposed in a previous proteomic study (Friso et al., 2004). Similarly to LQY1, the protein encoded by At2g34860 also only contains the cysteine rich zinc finger domain of DnaJ and not the other domains. The predicted mature protein possesses four conserved C-X-C-X-G-X-G zinc binding motifs, typical of DnaJ proteins, which are expected to bind a total of two zinc ions (Figure 15). In contrast to LQY1 however, the protein does not contain any

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hordeum</td>
<td>F2CR65</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 193</td>
</tr>
<tr>
<td>Ricinus</td>
<td>B9T070</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 189</td>
</tr>
<tr>
<td>Oryza</td>
<td>Q6YUA8</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 187</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>O64750</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 186</td>
</tr>
<tr>
<td>Populus</td>
<td>B9GLJ5</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 183</td>
</tr>
<tr>
<td>Vitis</td>
<td>F6GUF6</td>
<td>NNKGLLRRPGAKELLDKMYNGRILPGS 183</td>
</tr>
</tbody>
</table>

Figure 15. Multiple sequence alignment of the zinc finger DnaJ-domain protein At2g34860 (O6YUA8) and its orthologs from other plant species. The suggested twin arginine motif for translocation across the thylakoid membrane via the Tat-pathway is denoted in bold while the putative hydrophobic core region of the signal peptide is underlined. The conserved cysteine and glycine residues of the C-X-C-X-G-X-G zinc binding motif are highlighted in yellow and red respectively. The alignment was performed using Clustal O (1.1.0) with default parameter settings.
predicted transmembrane domains. Two lines of evidence, the putative Tat-pathway signal peptide and the co-expression with lumenal pentapeptide proteins, suggests that the protein encoded by At2g34860 is a lumen localized protein, and may represent a novel soluble component of the lumenal thiol-disulphide exchange regulation system. Which lumen- and thylakoid bound proteins it may interact with, and under which conditions however remains to be determined. Taken together with the recent discoveries made by other groups, this new finding illustrates that there probably are several more components of the thylakoid and lumenal redox regulation networks still to be discovered, making them much more complex than initially believed.

The growing number of lumenal proteins involved in the repair and assembly of PSII, including HCF136 (Meurer et al., 1998; Plucken et al., 2002), FKBP20-2 (Lima et al., 2006), PPL1 (Ishihara et al., 2007), CYP38 (Fu et al., 2007), TLP18.3 (Sirpio et al., 2007) and Psb27 (Liu et al., 2011), together with co-expression analysis, implied that also the pentapeptide repeat proteins may play related roles. Our finding that TL15 and TL20.3 exhibit folding chaperone activity towards a chemically denatured model substrate further supports this hypothesis. An unexpected finding however was that the reduced dithiol forms of the proteins showed increased chaperone activity, both by reduction/alkylation, and by site directed mutagenesis. As discussed above, this is in contrast to other redox-regulated lumen proteins such as FKBP13 and VDE, which are oxidatively activated, but it is however in line with peroxiredoxin Q, which is activated by disulphide reduction. One hypothesis is that the pentapeptide proteins may act as folding chaperones, assisting in the assembly of components of thylakoid membrane complexes on the lumen side, or may function in folding important lumen proteins imported via the Sec pathway, such as PsbO and plastocyanin. The current data is not strong enough to support such a hypothesis though and further experimentation is needed. To confirm and better understand the redox-regulated chaperone activity of the pentapeptide repeat proteins, novel assays are being developed, using recombinant lumenal substrates, rather than un-physiological model substrates. Also characterization using reverse genetics is a necessary next step, in order to better understand their function in relation to other processes in the chloroplast.

In conclusion, the chloroplast lumen contains a distinct proteome, which after its initial characterization has received increasing attention. While a growing number of proteins localized there are involved in the assembly and maintenance of PS II (Plucken et al., 2002; Chen et al., 2006; Lima et al., 2006; Fu et al., 2007; Ishihara et al., 2007; Sirpio et al., 2007), new distinct roles for lumen proteins, for example as components of the lumenal part of the NDH complex (Ishihara et al., 2007; Peng et al., 2009; Sirpio et al., 2009; Suorsa et al., 2010; Yabuta et al., 2010), or in protection against
oxidative stress (Levesque-Tremblay et al., 2009), are continuously being discovered. Also in this thesis a possible novel function for the pentapeptide repeat proteins as folding chaperones is proposed (Paper IV), a function which maybe also can be found among their cyanobacterial homologs. A detailed co-expression analysis approach may provide further clues to the functions of lumen proteins, and also to their interplay with other processes in the chloroplast. In addition it may aid in the identification of novel lumen proteins, such as the DnaJ domain protein encoded by At2g34680. A highly interesting aspect of the lumen proteome is its apparent regulation via dithiol/disulphide exchange reactions, mediated by redox proteins such as thioredoxins, thiol oxidoreductases and protein disulphide isomerases. The strong representation of in vitro targets of thioredoxin among the known lumen proteins, identified in Paper II of this thesis, emphasizes the importance such pathways may play in regulating the function and turnover of lumen proteins, in photosynthesis, in the protection against high light and in other processes. This will continue to be an exciting field of research, especially trying to understand the environmental signals and conditions controlling the redox state of the lumen and of individual lumen proteins.
Future perspectives

There are many aspects of the work performed in this thesis which I believe are interesting for future research, both in the context of photosynthesis and in increasing our understanding of thiol redox-regulation in general. To emphasize just a few specific questions of personal interest:

- The further identification and characterization of novel redox proteins in the lumen and thylakoid membrane, such as the putative disulphide isomerase encoded by At2g34860, which I identified by co-expression analysis.

- To better understand the thiol redox state of the lumen space under different environmental conditions, using a redox active form of GFP, imported into the lumen using an established luminal targeting sequence. This type of analysis has for example been performed in the endoplasmatic reticulum of mammalian cells, enabling real-time monitoring of redox changes (van Lith et al., 2011).

- The structure determination of the reduced form of TL15 would provide a structural basis to understanding the possible redox controlled chaperone activity of the pentapeptide repeat proteins.

- Solving and analyzing the structure of the PPD6 protein would provide the first structural insights into the apparent functional divergence of the PsbP-domain proteins, and into the possible regulatory role of its unique disulphide bond.
Acknowledgements

I have been guided, helped and accompanied by many people during the time of my work on this thesis. Among many others, I would especially like to thank:

Prof. Wolfgang Schröder, my supervisor and friend during these years. I still remember, like yesterday, just after my last exam, when we spoke about the possibility of a Phd studentship in your group. Had I the choice again, it would be the same; there is no other supervisor I would rather choose. You have guided and supported me, much more than you believe, not only in experimental science but also in so many other aspects. The Eye of Sauron now awaits us!

Dr. Thomas Kieselbach, my co-supervisor and mentor in all things. From practical labwork, to writing and project planning, your skills and knowledge as a scientist continue to amaze me, and you have taught me immeasureable amounts. In all my sidetracks and initiatives, you have always supported me with optimism and constructive suggestions. Without your ideas and inspiring project, this thesis would not be.

Prof. Christiane Funk, my co-supervisor and tree-climbing apprentice. You may not have needed to spend much of your time with my supervision, but your constant optimism and happy face has inspired me. It is what gives our place of work the fantastic environment it has. The FC will be finished. Eventually.

Former and current members of the Wolfman group. Jose Garcia, for helping me in the lab, teaching me history and generally being a great lab- and office mate. Liudmilla Shiryaeva, for companionship and good times in the lab. Irene Granlund for our friendship, great collaborations, patiently answering all my questions, teaching me so many practical aspects of proteomics, and the Christmas tree safaris. Yogesh Mishra for conversation, scientific discussions, great Indian food, constant enthusiasm, and for our many great collaboratory projects. I may not have bubbled with AhpC optimism from the start, but your enthusiasm and drive has brought it all the way through. My partner in gambling and protein purification, Patrik Storm. Although you haven’t made me rich yet, you have taught me and helped me with so many things throughout these years. Your company for coffee will be sorely missed when you move back down to the swamp. And last but not least, Lotta von Sydow, for being my masters student and letting me practice my supervision skills on you. Your company in the lab and under bridges has been fantastic. I may have been erratic at times, but you somehow always understood what I was trying to say, and took our projects forward.
To my collaborators in Lund, Warwick and Sevilla, who helped with important aspects of the work in this thesis, Prof. Hans-Erik Åkerlund, Dr. Colin Robinson, Dr. Marika Lindahl and the members of their respective groups.

Elisabeth Sauer-Eriksson and Stefan Jansson for having time to sit in on my evaluation meetings, always doing so with great enthusiasm and good advice.

To professors Stefan Jansson and Choon-Hwan Lee, it was in your labs my scientific career started. It was there I discovered the fantastic field of photosynthesis research, and I am very grateful for the time I had there and where it has led me.

To Lars Backman, for all issues regarding both teaching and biochemistry you always knew the answer and always were available to help me. As you know, I very much enjoyed and learned from the time I spent teaching and would love to do so more in the future.

Past members of the Funk and Backman groups: Adriana, Maribel, Irma, Miguel, Erik, Barbara, Helder and Harald, for great company in the corridor. And of course the hard working but still ever smiling current members, Otilia, Giada, Susan and Raik. My special thoughts also go to Tania Tibiletti, my apprentice in haggling, for our collaborations, discussions and great travelling company in China. For teaching me Spanish phrases, which are seemingly important for lab work (although not suitable for translation in this thesis), I give extra thanks again to Helder and Jose.

Marcus Wallgren, for being a great office mate during the first years and a good friend ever since, always there for a short chat or discussion, scientific or whatever. The same goes to Jörgen Åden, partner in taking conversations to a lower level, gambling and work. Ximena Aguilar, for all those conversations and good company over lunch. And of course to all other members of Gerhards, Magnus and Pernillas groups, for your company at coffee, lunch and in our corridors.

Special thanks goes to Uwe Sauer, for our collaborations and for teaching and helping me with all issues of the exciting but difficult field of protein crystallography. Also to Erik Lundberg, Christin Grundström and Elisabeth Sauer-Eriksson for all their assistance and suggestions.

The Messinger and Strand groups for interesting group seminars, and especially to Håkan Nilsson for having someone to share the experiences of fathership with.
To my family up north, **my mother, my father, Hans**, my little brother **Torbjörn** and to my big brother **Robert** and his family, for all the indirect support they have given me from afar. This also goes to my relatives in Scotland and my second family in Stockholm south. To my friend **Christian Poromaa** for all things from IKSU to football matches to just company. **Pernilla Linden** for friendship, morning coffee and all those trips to the mountains and other places. Finally my partner in life and eternal love, **Matilda Rentoft**, for her everlasting love and support, and for bringing into this world our own angel, **Alexandra** – thank you for being in my life.

As my last words I shall convey the best and most important advice I have received during the time of this thesis work. At almost the same time, just after my beginning this work, my supervisor Wolfgang and my mother Sheila both told me: “Learn to get things done”. Simple as this short piece of advice may seem, it is not always so in practice, and I will always continue to carry it with me.

The work in this thesis was supported by personal research grants from the SJCKMS foundation and Stiftelsen Lars Hiertas Minne, and travel grants from Kungliga Skogs- och Lantbruksakademien.
References


Anbudurai PR, Mor TS, Ohad I, Shestakov SV, Pakrasi HB (1994) The Ctpa Gene Encodes the C-Terminal Processing Protease for the D1 Protein of the Photosystem-II Reaction-Center Complex. Proceedings of the National Academy of Sciences of the United States of America 91: 8082-8086


57


Lindahl M, Florencio FJ (2003) Thioredoxin-linked processes in cyanobacteria are as numerous as in chloroplasts, but targets are different. Proceedings of the National Academy of Sciences of the United States of America 100: 16107-16112


Motohashi K, Hisabotri T (2010) CcdA is a thylakoid membrane protein required for the transfer of reducing equivalents from stroma to thylakoid lumen in the higher plant chloroplast. Antioxid Redox Signal 13: 1169-1176


Yamamoto HY, Kamite L (1972) Effects of Dithiothreitol on Violaxanthin Deepoxidation and Absorbance Changes in 500-Nm Region. Biochimica Et Biophysica Acta 267: 538-


Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-
O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by
transcriptome coexpression analysis and reverse genetics. J Biol Chem 282: 14932-
14941

(2008) Sorting signals, N-terminal modifications and abundance of the chloroplast