# Table of contents

**Abbreviations** ................................................................................................. 2

**Introduction** ................................................................................................. 3
  Global perspective ....................................................................................... 3
  Proteome and proteomics ............................................................................ 4
  The chloroplast ............................................................................................. 4
    Chloroplast ............................................................................................... 4
    Stroma ....................................................................................................... 6
    Thylakoid membrane ................................................................................. 6
    Thylakoid lumen ....................................................................................... 7
  Photosynthetic electron transport ................................................................. 8
  Light signalling, circadian rhythm and acclimation ........................................ 10
  Reactive oxygen species ............................................................................. 12

**Key techniques** ........................................................................................... 13
  Two Dimensional Fluorescence difference gel electrophoresis ..................... 13
  CyDye™ labelling ....................................................................................... 14
  Gel imaging and analysis with DeCyder™ .................................................. 14
  Spot picking ............................................................................................... 16
  Clustering of mass spectrometry spectra .................................................... 17

**Results** ........................................................................................................ 18

  **Paper I** .................................................................................................. 18
    The chloroplast stroma proteome from Arabidopsis and Spinach ........................................................................... 18
    Clustering of peptide mass fingerprint data ............................................... 19
    Protein modifications: interesting but problematic ..................................... 20
    Search for PTMs with “peaks-in-common” .................................................. 20
    LC MS/MS improves identification of complex samples .............................. 21

  **Paper II** .................................................................................................. 21
    A faster lumen preparation ....................................................................... 21
    Thylakoid lumen proteome in dark versus light adapted plants .................. 22
    Lumen proteins attachment to the thylakoid membrane ............................. 22
    Light regulation in plant ......................................................................... 23

  **Paper III** ................................................................................................ 24
    Isoforms and homologues in plants ............................................................ 24
    Homologous isoforms of plastocyanin ....................................................... 24
    Movement of plastocyanin in the thylakoid lumen ..................................... 26
    Cyt c6A and plastocyanin .......................................................................... 26
    Amount of plastocyanin .......................................................................... 26
    Photosynthesis performance in the mutants ................................................. 27
    The role of copper .................................................................................... 27
    The role of plastocyanin isoforms and amount ......................................... 28

  **Paper IV** ................................................................................................ 28
    Thylakoid lumen protein of 29kDa ............................................................. 28

**Summary** ...................................................................................................... 30

**Perspective** ................................................................................................. 31

**Acknowledgements** ................................................................................... 33

**References** ................................................................................................. 35
**Abbreviations**

APX       Ascorbate peroxidase  
Cyt       Cytochrome  
2D-PAGE   Two dimensional polyacrylamide gel electrophoresis  
DIGE      Difference gel electrophoresis  
DNA       Deoxyribonucleic acid  
IEF       Isoelectric focusing  
MALDI     Matrix-assisted laser desorption/ionization  
MS        Mass spectrometry  
PCp       Plasrocyanin protein  
PLAS1     Plastocyanin 1, minor protein (At1g76000)  
PLAS2     Plastocyanin 2, major protein (At1g20340)  
PMF       Peptide Mass Fingerprints  
PSI       Photosystem I  
PSII      Photosystem II  
PTM       Post-translational modification  
ROS       Reactive oxygen species  
SDS       Sodium Dodecyl Sulphate  
TL29      Thylakoid luminal 29kDa protein (At4g09010)  
TOF       Time of Flight  
WT        Wild-type
Introduction

Global perspective

Year 1771 Joseph Priesley made an important discovery, plants allowed mice to survive when they were put together in airtight glass containers, figure 1. The conclusion he drew was that plants have an ability to clean the air that animals and humans have polluted (Priesley 1776). Even if he actually did not fully understand the experiment and that oxygen was involved, this experiment is often referred to as the start of photosynthesis research. Today we know that the overall equation of photosynthesis is:

\[ \text{(1) CO}_2 + 2\text{H}_2\text{O} \leftrightarrow \text{O}_2 + (\text{CH}_2\text{O})_n + \text{H}_2\text{O} \]

In this process solar energy is converted into chemical energy. Inorganic carbon in the form of CO\textsubscript{2} is converted to organic carbon as starch, under oxidation of water to oxygen and protons. Water is on both sides of equation 1, to show that it is not the same water. This process can be performed by eukaryotic organisms such as plants, trees and algae but also by prokaryotic such as cyanobacteria. This process that looks quite simple and straightforward, has had a spectacular impact on our earth. The by-product, oxygen, created oxygenic atmosphere and an ozone layer so that ocean evolved life could crawl up on land. Beside this crucial point for life on earth, photosynthesis gives us direct or indirect the food we eat, fuel, many different construction materials and much more. It is almost impossible to find something that does not originate from photosynthesis.

Today global warming and CO\textsubscript{2} release are discussed daily and headlined as the greatest threat to our earth. The problem arises from infrared-absorbing trace gases such as carbon dioxide and methane in the atmosphere. The gases absorb infrared radiation that normally would be reflected out into space, which leads to increased temperatures on earth. Carbon dioxide is responsible for about half of the atmospheric heat retained by trace gases and is produced primarily by the burning of fossil fuels (Manahan et al. 2000). The only known process in which the carbon-dioxide is reduced is the process of photosynthesis. Therefore an active production of green plants as forest
(trees) is an effective move towards decreasing atmospheric carbon dioxide levels, and also for energy production. As an example one hectare can in one year produce 8 tons of dry wood which corresponds to 3.2 m³ oil (Christersson et al. 2005). Thus understanding of the photosynthesis process, down to the molecular level is an important field of research, both because it is the driving force of life on our planet and also because it can solve our energy problems.

**Proteome and proteomics**

Words with the suffix 'ome' define the varied population and sub-population in a cell and the same words that end in 'omic' are associated with the corresponding research in the area. These suffixes are now widely used in the field of science. One of the first 'ome' words was genome which was built from the words gene and chromosome and more precisely means the complete genetic sequence on one set of chromosomes of an organism (Lederberg and McCray 2001). To describe the set of proteins expressed by the genome the proteome term was coined, in the mid-1990's. In the original definition it meant the proteins that are expressed in a cell or a defined compartment in the cell at a specific time point. Today the proteomics term also includes determining of the location, modification, interactions, activities and functions of proteins (Greenbaum et al. 2001, Tyers and Mann 2003). It also includes a special set of technology used, such as two dimensional gel electrophoresis and various types of mass spectrometry and HPLC. As most protein identifications are made by peptide mass fingerprint (PMF) identification, the effort which in recent years has been put into performing large genome sequencing projects, has formed the base for the development of proteomics.

**The chloroplast**

**Chloroplast**

According to the symbiosis theory the plant chloroplast is believed to have the same ancestor as cyanobacteria, (Raymond and Blankenship 2004). The chloroplast has it’s own circular DNA, with roughly 100 genes, the rest is encoded by nuclear DNA, synthesised in the cytosol and transported to the chloroplast organelle, (Jarvis 2004, Martin et al. 2002). Of the whole Arabidopsis genome, containing around 28,000 genes, 3100-4200 genes were predicted to produce chloroplast located proteins, (Kieselbach and Schröder 2003, Kleffmann et al. 2004, Sun et al. 2004). A prediction on the TAIR home page (http://www.arabidopsis.org) with the search keyword “chloroplast” revealed 2696 chloroplast located genes (2257 loci) in Arabidopsis.
Most cells in the leaves, except for the outer cells (epidermis), contain the chloroplast organelle where photosynthesis is performed, figure 2, (Nelson 2005, Karp 2002). Plant chloroplasts have a disk-shaped structure with a length up to 10 µm and a width of 3-4 µm. The number of chloroplasts in mesophyll cells ranges from 2 to 120, (Koniger et al. 2008). The chloroplast consists of three membranes; the envelope, the boundary to the surrounding plant cell, consisting of an external and an internal membrane. The envelope encloses the stroma and the thylakoid membrane. These membranes enclose separate areas, which do not seem to be in direct contact with each other, see figure 3 (Dekker and Boekema 2005).
Stroma

The stroma compartment is the location for the Calvin cycle, which is a metabolic pathway where carbon enters in the form of CO$_2$ (Karp 2002). The chloroplast stroma is considered to mainly contain hydrophilic proteins, but because the stroma compartment has been shown to have a high protein concentration (400 mg/mL, out of which 250 mg/mL is ribulose-bisphosphate carboxylase (Rubisco)), the question is how soluble this compartment actually is. Thus a tight packing of stroma proteins are assumed, which probably reduces the random movement of the proteins in the stroma. In fact large dense clusters of stroma proteins can be observed with cryoscanning electron microscopy, which can form channels for solvent movement, (Pessarakli 2005). About 75% of the stroma proteome consists of proteins involved in the oxidative pentose phosphate pathway, glycolysis and the Calvin cycle. Proteins involved in chloroplast protein synthesis, biogenesis and lipid metabolism were only represented by 10% of all identified proteins. Nitrogen assimilation was represented by 5-7% and fatty acid and amino acids pathways were represented by less than 1% (Peltier et al. 2006).

Thylakoid membrane

Thylakoid membranes are divided into grana and stroma structures, see figure 3 and 4. Grana have a cylindrical structure with 10-20 stacks, 300-600 nm in diameter, that are linked to each other through stroma lamellae which are several hundred nanometres long. The photosynthetic light reaction, were ATP and NADPH is produced by absorbed light, takes place in this membrane. For this purpose the thylakoid membranes contains four supra-protein-complexes: photosystems one (PSI) and two (PSII), together with their light-harvesting complex protein (LHCP), the cytochrome (Cyt) b$_6$/f complex and the ATP synthase. These complexes are working together in series to convert light energy to chemical energy in the form of ATP and NADPH, figure 6, (Karp 2002, Nelson and Cox 2005). About 85% of the PSII are located in the appressed domain of grana, figure 4. The remaining PSII are found in the stroma lamellae and end membranes. PSI with the larger antenna, involved in linear electron transport, are located in the grana margins while PSI with less antenna are located in the stroma lamellae, (Albertsson and Andreasson 2004). Cyt b$_6$/f is distributed more evenly in the thylakoid membrane with lateral movement depending on light conditions and photosynthetic electron flow (Vallon et al. 1991). The proton-motive force driven ATP-synthase, with its proton-driven motor on the stromal side of the thylakoid membrane, does not have enough space to be located in the appressed grana domain and thus is mainly located in the stroma lamellae (Nelson and Cox 2005).
Thylakoid membranes have a larger content of galacto-containing glykolipids than of phospholipids. Double bonds in glykolipids give the lipid layer of the thylakoid membranes a high mobility, which means that thylakoid membranes can easily be rearranged and the proteins can easily move around in the membrane length direction by lateral diffusion, which has a key role in photosynthesis, (Borodich et al. 2003, Karp 2002).

**Thylakoid lumen**

Since the majority of proteins needed in the chloroplast are synthesised in the cytosol, they need a transport system through the envelope and through the thylakoid membrane. This transport system is especially important for lumen proteins as all lumen proteins found up to date are synthesised in the cytosol. The protein transport mechanism in chloroplasts is a mix of both prokaryotic and eukaryotic export pathways and import routes, for details on the transport see review (Jarvis and Robinson 2004).

The number of experimentally found thylakoid lumen proteins has increased since the beginning of the 1960’s when plastocyanin was the first lumen protein to be discovered, (Katoh et al. 1961). The amount of thylakoid lumen proteins is believed to be around 80 proteins according to Kieselbach and Schröder, 2003. Today at least 70 different lumen proteins from Arabidopsis have been experimentally identified, (Schubert et al. 2002, Gupta et al. 2002, Peltier et al. 2002 and 2004, Goulas et al. 2006, Petersson et al. 2006). A search on the TAIR home page (http://www.arabidopsis.org) using the keyword “thylakoid lumen” results in 84 distinct gene models and 70 loci matches. Of these 84 genes, 31 were not experimental identified as thylakoid lumen proteins, meanwhile an additional 20 proteins that were experimental identified thylakoid lumen proteins were not among these 84 genes. This shows that there are still thylakoid lumen proteins to be identified. As much as 30% of the predicted lumen proteins have an unknown function, 22% are closely related with PSII and 19% are immunophillins. Surprisingly as many as 29% have a unique function, figure 5.

*Figure 4. A model of the grana structure with a helical arrangement and different grana diameter in the stack. The grana structure is divided in different parts following fractionation with for example digitonin.*

Figure 5. Classification of predicted lumen proteins; 30% unknown proteins, 22% related to PSII, 19% putative immunophilins and 9% proteases. The remaining 20% are more or less unique proteins.

The compartment of the lumen space is really narrow and the proteins are therefore always close to each other and the thylakoid membrane. They can therefore also often be peripheral with different degree of attachment to the membrane, (Paper III). The location of the different proteins within the lumen compartment varies, and often depends on what function the protein has. Lumen proteins found in appressed grana are close to the PSII complex and are therefore often believed to be related with this complex.

Photosynthetic electron transport

Light excites PSII's reaction centre chlorophylls, P$_{680}$, which become oxidised to P$_{680}^{+}$, which has the highest redox potential observed in a biological system. The high potential enables water splitting and electron transfer through Mn-clusters and a reactive tyrosine Yz. Electrons from the water splitting then reduce the oxidised P$_{680}^{+}$ back to P$_{680}$. To produce one molecule of molecular oxygen four light quanta to oxidise P$_{680}^{+}$ and two water molecules to reduce P$_{680}^{+}$ back to P$_{680}$ are needed. The released electrons from P$_{680}$ are transported through an accessory chlorophyll and a pheophytin molecule to a bound QA. A mobile QB is reduced, taking up two electrons (QB$^{2-}$) and
also two protons from the stroma to form Q$_b$H$_2$. Q$_b$H$_2$ is released in the lipid bilayer of the thylakoid membrane were oxidised quinone (PQ) are reduced (PQH$_2$). Translocation of electrons through the thylakoid membrane also results in transport of protons from the stroma side of the thylakoid to the lumen side. The electrons reach the Cyt $b_6/f$ complex and Cyt $f$ then transfers them to the copper binding electron carrier protein, plastocyanin (PCp) in the thylakoid lumen. In Arabidopsis there are two isoforms of PCp's with high similarity, PLAS1 (the minor isoform) and PLAS2 (the major isoform) (paper III). PSI, containing P$_{700}$, is similarly as PSII oxidised by light. On the lumen side of PSI, PCp docks to PsaF and reduces the oxidised P$_{700}$ in PSI. Electrons pass through PSI to the stromal side of the thylakoid, where ferredoxin (Fd) is reduced. Ferredoxin-NADP$^+$-reductase (FNR) transfers electrons from reduced Fd to NADP$^+$ to form NADPH (Nelson and Ben-Shem 2004, Renger 2008).

**Figure 6.** Schematic description of Photosystem II, cytochrome $b_6/f$, photosystem I, the ATPsynthase complex and the principal electron transport and proton pathway. Linear electron flow is indicated by black arrows and cyclic electron flow with dark blue dashed arrows. The two isoforms of plastocyanin, marked in blue, transfer electrons from cytochrome $b_6/f$ to PSI in both linear and cyclic electron transport.
The electron transport chain generates a proton gradient which is used by CF₀-CF₁ ATP synthases for the formation of ATP, the power molecule of the organism, (Choquet and Vallon 2000). When there is too low amount of NADP⁺ or if the plant needs to vary the ratio of NADPH to ATP formed in the light, the plant uses an alternative pathway where the electrons are recycled from PSI to Cyt b₆/f. Fd donates electron back to the Cyt b₆/f complex, which through PCp can reduce P₇₀₀⁺⁺ when the plant is illuminated. The electrons propeller around PSI in a cyclic manner, without net formation of NADPH or O₂. However under these condition a pH gradient is formed and thereby ATP can be synthesised, figure 6, (Nelson and Ben-Shem 2004, Renger 2008).

**Light signalling, circadian rhythm and acclimation**

In plants light is sensed by photoreceptors such as phytochomes and cryptochromes as well as by chloroplast redox regulation. The light signal can then induce changes in gene expression, which in turn regulates biological processes. It has been shown that as much as 30% of genes expressed in Arabidopsis leaves show some degree of diurnal regulation (Bläsing et al. 2005). The interplay between the different light signalling pathways, sugar metabolism and other processes is very complex and in addition to this the plants uses a programmed response to control gene regulation known as the circadian clock.

Jean jacques d'Ortuos the Marian was a French researcher, and in the year 1729 he demonstrated that plants showed a daily sheet movement for several days, even in the dark. He had thus shown that the circadian operating income was not dependent on the daily light and dark cycle even if it was synchronised with this, (Coleman 1986). The clock runs on a period of approximately 24 hours in constant light, dark or carbon dioxide limited conditions. The clock output is highest around noon when the light is strongest (in the case of natural conditions) and lowest in middle of the night, (Hennessey et al. 1993) figure 7.

![Figure 7. Circadian rhythm shown as sinus waves which often have a period of approximately 24 hours.](image)
Biological activities that are controlled by the circadian clock are; opening of the stomata in the day, the closure of bloom leaves at night and control of gene expression, (Niinuma et al. 2007, McClung 2006). Microarray analysis of more than 8000 Arabidopsis genes found that 6% of the genes are circadian regulated. These circadian regulated genes are involved in the plant response to light pathway, other important metabolic pathways such as carbon, nitrogen and sulphur metabolism and also transcription factors and protein kinases, (Harmer et al. 2000, Schaffer et al. 2001). Circadian rhythm ensures that that plants grown under natural conditions maintain maximise rates of photosynthesis near noon and minimise rates near midnight, (Hennessey et al. 1993). Plants have the ability to match the light-dark cycle to circadian rhythm in order to increase survival, and also grow better by holding more carbon, (Dodd et al. 2005).

Since the plant does not have the ability to move, it has developed a number of mechanisms to adapt to the prevailing environment, and especially differences in light changes. In addition to circadian rhythm, which provides early responses to light changes, the plant responds directly by re-organising the structure and function of the chloroplast components, so-called short-term responses. Changes in PSII effectiveness, (Horton et al. 1996) and RuBisCo activity, (Salvucci and Ogren 1996) are examples of short-term responses made over a period of seconds to minutes. Long-term responses to the light change, or so-called acclimation, are changes in the amount of chloroplasts and changes in composition and function of the photosynthesis apparatus. Depending on whether the plants are exposed to low or high light, their responses differ. During the acclimation to low light the amount of PSI are increased compared with other light levels. This may be because the plant needs to store more ATP when they grow very slowly and are close to their light compensation point. Plants have a distinct approach to acclimate for growth in extremely high or low light, and these strategies involve not only changes in thylakoid protein levels, but also involve major structural and functional changes in one or both photosystem, (Bailey et al. 2001).
Reactive oxygen species

In the thylakoid lumen reactive oxygen species (ROS) can “leak” out from the oxidising (water-splitting) side of PSII, figure 8 (McKersie 1996, Apel and Hirt 2004). ROS are small molecules that are highly reactive due to the presence of unpaired valence shell electrons. ROS cause oxidative damage to molecules such as lipids, proteins and DNA. The damage to DNA can then cause mutations. Beside oxidative damage ROS also control and regulate different biological processes such as growth, the cell cycle (Gapper and Dolan 2006), programmed cell death, hormone signalling, biotic and abiotic stress responses and development, (Shao et al. 2008, Bailey-Serres and Mittler 2006, Matés and Sánchez-Jiménez 1999).

Figure 8. Production of ROS particles in the chloroplast stroma and thylakoid lumen. Response and regulation control mediated by the ROS particles.
This study

In this work I have studied two proteomes, the stroma and the thylakoid lumen. Both of these have been analysed by 2D-gels and the separated spots have been identified by mass spectrometry. I have then performed functional genomic analysis of two proteins identified in the thylakoid lumen. The questions that I have addressed are:

❖ Are there any distinct differences between the stroma proteomes of two different plants? Could similarities be used to facilitate increased protein identification? Could analyses of peptide mass fingerprints help in identifying protein post-translational modifications?

❖ How dynamic is the thylakoid lumen proteome? Are there any differences in the lumen proteome between dark and light acclimated plants? How well do proteome changes correlate to gene expression?

❖ A large part of the Arabidopsis genes are present in more than one copy, as is the case for plastocyanin. Is there a functional aspect of this duplication?

❖ TL29 is a lumen located ascorbate peroxidase homolog. What function does it have?

Some of the techniques used in this thesis are crucial for the proteomic analysis and thus described below.

Key techniques

Two Dimensional Fluorescence difference gel electrophoresis

Two Dimensional Fluorescence difference gel electrophoresis (DIGE) is a method that is used to measure up and down regulation of protein spots on a 2D gel. Instead of staining the gels after separation by 2D-PAGE, a fluorescence probe is covalently added to the proteins before they are separated. One of the most commonly used methods is minimal labelling, where a fluorescence molecule is covalently bound to lysine residues. Up to three different minimal fluorescent Cyanine dyes (CyDye™) are used to label three different samples which are mixed together and separated on the same gel, thereby exposing the samples to the same chemical environments and electrophoretic conditions. One of the dyes is used as an internal standard, containing a pool of each sample used in the experiment in order to ensure that all the proteins from the different samples are represented. The internal standard is used in every gel in the experiment, significantly reducing gel to gel variation.
of protein ratio measurements. Ünlü et al (1997) was one of the first to describe this technology (Ünlü et al. 1997), and many have followed with improvements, (Gharbi et al. 2002, Herbert et al. 2001, Lilley et al. 2001, Alban et al. 2003).

**CyDye™ labelling**

The samples are solubilised in an appropriate buffer that should not interfere with the colouring. To ensure that approximately only 3-5% of the available lysine residues and that only a single lysine per protein molecule is labelled, the quantity of dyes added to the sample is limited and the reaction is controlled by time and temperature. At the end of the reaction lysine is added in surplus, quenching any unused colour.

After the proteins have been labelled the samples are further solubilised, to increase the sample volume and to change the pH to the buffer suited for isoelectric focusing (IEF). In all tubes one third of the total protein amount should contain the mixed standard sample, preferably labelled with CyDye™ 2. The other two thirds are comprised of the different samples, each labelled with one of two different colours, CyDye™ 3 and 5. A sufficient number of biological replicates, preferably four or more, should be made to obtain good statistics, (Karp and Lilley 2007). Swapping of the two CyDye™’s in replicate samples is used to get around the dye specific differences in signal intensity and thereby get a better statistical analysis. Finally one tube contains a mixed sample (standard), and two samples. Each pooled sample is then separated by iso-electric focusing (IEF) and 2D-PAGE, see work flow in figure 9.
After separation on 2D-PAGE the gels are scanned using a laser scanner with an appropriate filter for the three colours. Because three colours are used, three images are obtained, each of them corresponding to one dye that is separated in exactly the same conditions. The sample mixes are the same in every gel, and therefore the image for the mixed sample (CyDye™ 2) should theoretically have the same spot patterns and the same intensity in all gels. In addition every spot that may occur in the various tests are included in the mixed sample.

DeCyder™ is one of the analysis programs for DIGE. The program matches the gels, normalises them between each other using the standard colour and calculates up- and down-regulation of spots by statistical analysis. It is then necessary for the user to judge a threshold for significant regulation of the spots in different samples. The DeCyder™ program consists of different parts, one is Difference In Gel analysis (DIA), were analysis of one particular gel is performed in order to identify up- and down-regulation for that specific gel. In the Biological Variation Analysis (BVA) program (figure 10), several gels are matched to each other and normalised. Here the statistical comparison between the samples is made. DeCyder™ also contains a batch processor were the gels are loaded and automatic normalisation and matching are performed, thereby decreasing manual work for the user. The statistical program in DeCyder™ uses the t-test, false discovery rate (FDR) and one and two way ANOVA for statistical analysis. The average ratio is shown as positive numbers for up-regulated proteins and negative numbers for down-regulated proteins. The average ratio is calculated for each spot as the average of standard abundance ratios. Standard abundance is the volume ratio divided by the normalised volume of the internal standard. The volume ratio is normalised background corrected volumes calculated as expressed ratios of Cy3/Cy5, Cy3/Cy2 and Cy5/Cy2.

![Figure 10. Screenshot from the DeCyder™ program BVA, with the four different working windows in protein table mode. The amounts of minor plastocyanin in the standard sample and in the sample from the mutant lacking the major form of plastocyanin are shown in the gel and 3D windows. The protein table presents the t-test results and a graph view shows standard log abundance of the minor plastocyanin spot in all different samples.](image)
Spot picking

Before spot picking after a DIGE experiment the gels are stained with another dye, for example coomassie dye or silver stain. Deep purple which is more sensitive than coomassie- and silver staining can be used with an automatic picking device enabling the picking of more spots. To use this device it is of great benefit if the gel is bound to one of the glass plates during gel casting, as this prevent the gel from swelling during staining which in turn facilitates the matching between the picked gel and analysis gel.

Because the fluorescence label is covalently bound to the lysine amino acid and it carries a +1 charge at neutral or acidic pH, the CyDye™ DIGE Fluors are matched to ensure that the pI of the proteins doesn’t change. The mass differences obtained upon coupling of CyDye™ are 434-464 Da. This gives a migration difference between unlabelled and labelled protein, therefore enforcing the user to post-stain the gel in order to ensure that the majority of unlabelled proteins are picked, which is necessary for obtaining sufficient protein for mass spectrometry (MS) identification. Migration difference is more significant for lower molecular weight proteins (GE Healthcare instructions RPK0272, 2008), Hrebicek et. al. (2007) also found that there is a risk that the wrong protein is picked if CyDye™ coordinates are used, especially in regions were there is a high density of proteins in the gel. I found also that the migration shift is more significant for lower molecular weight proteins and that there is a high degree of disparity between individual proteins, giving a considerable scattering of the data, figure 11.

![Figure 11](image)

**Figure 11.** The correlation between the migration difference, between coomassie stained spots and CyDye™ 2 labelled spots, (y-axes) and the ratio between the total migration distance of CyDye™ 2 labelled proteins and the front marker (x-axes).
Clustering of mass spectrometry spectra

Normally various types of clustering are used when comparing for instance plant genes, to understand their relationship and obtain better understanding of evolutilonal pathways. However, also PMF masses can be treated as if they were different species and be clustered. Large data sets of MALDI-MS mass spectra from protease-digested proteins can be clustered by hierarchical clustering using SPECLUST, figure 12, which is a web tool provided at http://bioinfo.thep.lu.se/speclust.html (Alm et al. 2006). PMF mass lists derived from spots excised from 2D-gel samples can be loaded in the web tool, as zipped .peaks files. SPECLUST then clusters mass lists which have the same masses together. The output is a dendrogram where the proteins that have similar PMFs cluster together, implying that these are proteins with similar amino acid sequence. Before the PMF mass lists are loaded into SPECLUST clustering the lists have to be “cleaned” from keratin, rests from pepsin used, matrix and other contaminating masses which are present in many mass lists. “Cleaning” of PMF mass lists is an important step, because spots with high amount of contamination will otherwise cluster together. The Peaks-in-common tool can be used to clean the PMF mass lists from contaminations as it identifies masses that are shared by a set of mass spectra.

In the Clustering window, metric, linkage and match score can be chosen. Metric selection is the use of method for measurement of the distance between the items to be clustered. Use of liberal metric, average linkage and a $\sigma$-value set to one dalton was found to work well for clustering of MALDI-TOF mass spectra, (Alm et al. 2006). The dendrogram, which is the outcome from clustering, is used for cluster identification. A cut-off value for the distance in the dendrogram is set for cluster identification. Mass spectra joined in nodes will be considered a cluster if the cut-off

![Figure 12. SPECLUST web tool window. A. Clustering window. B. Peaks-in-common window.](image)
value is higher than the node distance. For example in figure 13, if a cut-off at \(d<0.8\) is used there will be only one cluster from all three peaks, but if a cut-off at \(d<0.1-0.7\) is used only S523 and S408 will form a cluster. Each cluster that is derived from the clustering can be submitted to the peaks-in-common application for further analysis.

**Results**

**Paper I**

**The chloroplast stroma proteome from Arabidopsis and Spinach**


In my work (Paper I) I extracted chloroplast stroma from Arabidopsis and separated the proteins by 2D-PAGE using in the first dimension IEF with a pH gradient of 4-7. Different to native gels, that separate protein complexes, IEF separates proteins according to their charge and isoelectric point (pI). The second dimension, SDS PAGE, then separates according to protein size/mass. In total 232 spots were visible after the second dimension, of those I was able to identify 208 spots, corresponding to 116 different proteins, resulting in an average of 1.8 spots per protein (supplemental table 1S Paper I). Using a wider pH range probably had resulted in better protein coverage, comparable to the one by Peltier et al. (2005). Interesting to note is that the large subunit of RuBisCo was represented with 27 spots in the SDS-PAGE. The large number of protein spots containing RuBisCo suggests that this important protein undergoes many different post-translational modifications (PTMs), pointing to a very complex regulation. The large amount of RuBisCo spots could also be a result of degradation products of the protein in the stroma compartment.

To compare my data obtained in the model organism Arabidopsis to other plants, I also obtained the stromal proteome map from Spinach. As seen in paper I, I was able to identify 132...
spots in the stroma compartment of Spinach using MS, compared to 208 spots in Arabidopsis. Using MS/MS instead of MS on Spinach samples, the amount of identified spots was increased from 48 % to 61 %. Compared to Arabidopsis less contamination of luminal proteins was detected in the stroma fraction of Spinach. Leakage of the thylakoid lumen is known to occur, but Spinach chloroplasts seem to be more robust during biochemical preparations and are therefore a better choice for chloroplast subproteomes and biochemical analyses. The 132 spots in the Spinach stromal fraction represented 80 different proteins, the other spots were due to for example, PTMs or multiple isoforms. I was able to identify 19 photosynthesis-related protein spots in Arabidopsis and 18 in Spinach. Additionally 108 protein spots were identified in Arabidopsis - 104 in Spinach - as being involved in metabolic pathways. The coverage of enzymes involved in the Calvin cycle (carbon dioxide fixation) was excellent, both in the Arabidopsis and Spinach data sets. In total I could identify 90% of all picked spots using preparations from Arabidopsis, but only 48% of the spots containing Spinach proteins. As mentioned earlier the amount of identified proteins involved in photosynthesis and metabolic pathways were nearly same in both species, but there were large differences in the identification of proteins with “other functions”. The genome of Spinach is not sequenced therefore more conserved proteins – with high homology to proteins encoded in the Arabidopsis genome – were easier identified. To increase the protein identification rate in an organism without a sequenced genome I did PMF clustering.

Clustering of peptide mass fingerprint data

All my data derived from MALDI-TOF analyses of the chloroplast stromal proteome using Spinach and Arabidopsis was clustered using SPECLUST. Two dendrograms were built, one containing all data derived from Arabidopsis and Spinach and the second one only containing the data from Spinach. First the PMF mass lists were loaded into the Peaks-in-common subprogram (available at the SPECLUST web page). Peaks from the mass spectra which were detected in many of the mass lists were inspected manually using different cut-offs. These peaks often derived from contaminations like keratin, trypsin or from matrix-related peaks. The PMF mass list was then submitted to SPECLUST for clustering.

Clustering the PMF mass lists containing the combined data-set from Arabidopsis and Spinach did not significantly improve the protein identification, most likely due to the poor sequence identity. However, clustering of the data set only derived from the Spinach PMF mass lists resulted in an increased protein identification from 48% to 61%. Clustering improved the identification to the same extent as MS/MS, clustering therefore seems to be equal to MS/MS. Considering that
SPECLUST is easier, cheaper and faster to use than MS/MS, it could be called the “poor man’s” MS/MS tool. When SPECLUST was combined with MS/MS (paper I table 1), the total identification rate increased further to 69%, as the two methods generally identified different protein spots. Identification of proteins via clustering is dependent on: I) clustering of the protein of choice with other proteins and II) identification of other proteins in the same cluster, and therefore is limited. The use of MS/MS on the other hand is dependent on the generation of good spectra.

**Protein modifications: interesting but problematic**

An important key regulator for structure, function and interaction of proteins, is PTM. Amino acid residues are chemically modified by attachment of functional groups like acetate, phosphate, lipids or carbohydrates. In this way the chemical nature of the amino acid is changed, but also structural changes (disulphide bridges), or removal of amino acids are possible. On the UNIMOD web-page [http://www.unimod.org/modifications_list.php](http://www.unimod.org/modifications_list.php) there are as many as 861 different modifications listed. In the FindMod tool at the Expasy web-page [http://www.expasy.org/tools/findmod/](http://www.expasy.org/tools/findmod/) 71 PTMs are listed, including amino acid substitutions and modifications that are performed *in vivo* as well as *in vitro*, like e.g. carboxyamidomethylation of cysteine, which improves protein solubility and identification.

Many post-translational modifications alter the pI of a protein and/or induce minor changes to the molecular mass, which then can be detected by 2D-PAGE (Rabiloud 2002, Finne et al. 2008). However, one should keep in mind, that not only PTM induces the appearance of multiple spots in 2D-PAGE, also alternative splicing, which generates different mature mRNAs originating from one precursor, or the existence of “multiple gene copies” have the same effect. The situation is further complicated by the fact that many plants are polyploid, with small mutations on the chromosomes which might exchange single amino acids in a protein, (allelic variation). This exchange can cause a different pI and result in multiple locations of the protein in the 2D gel. Further on there might be non-biological reasons for multiple spots, the degradation of urea in the sample buffer for example can cause carbamylations (McCarthy et al. 2003), which can be avoided by proper sample handling.

**Search for PTMs with “peaks-in-common”**

Using “Peaks-in-common” separately for each cluster obtained in SPECLUST, identical masses that occur in several spots can be filtered out. Adding the average mass-list and the amino-acid sequence of identified proteins into the program “FindMod” we received a list with potential PTMs and potential single amino acid substitutions. However, due to the high number of possible
PTMs it might be difficult to distinguish which one is most likely. Masses that are not identical within a cluster could indicate modified peptides, but could also be due to contaminations from one or more proteins co-migrating in the same gel spot.

I used “peaks-in-common” for one cluster with four spots (paper I). Two data sets for each spot were loaded into “peaks-in-common” and when comparing the mass lists I could detect differences between the spots (paper I, table 4). They were identified as the small subunit of RuBisCo and were located in the 2D-gel at pH 6 (two spots) and pH 7 (another two spots) with a molecular weight of ~15 kDa. MS/MS analysis identified besides other PTMs an amino-acid substitution, explaining the different pIs of the proteins.

Following the work flow presented in paper I (figure 6) in a more extensive manner, the probability to identify in the same cluster PTMs leading to different gel spots will increase.

**LC MS/MS improves identification of complex samples**

One cluster (cluster 78 in paper I) was located in an area with a large amount of spots, which had not been identified by MS. The cluster (78) was closely located to another cluster (79), identified as acetohydroxyacid isomeroreductase. The proteins of cluster (78) were separated by liquid chromatography (LC) and analysed by MALDI-TOF-TOF. They were found to contain both acetohydroxyacid isomeroreductase and glutathione reductase, identification of the proteins was only possible because of the improvement due to LC. Even though LC-MS/MS is rather time consuming when separating the peptides, it is convenient in handling as the samples are automatically separated and spotted on the MALDI target plate.

**Paper II**

**A faster lumen preparation**

The thylakoid membrane, containing all the proteins needed for the photosynthetic light reaction, encloses a space (lumen) which is full of proteins. Light is the key regulatory factor in photosynthesis, the question therefore arises about the effect of light on the luminal proteome. To be able to compare leaves of dark or light adapted plants I had to develop a thylakoid lumen preparation that was faster than the usual one, which takes around 7 hours, to reduce the risk of changes in protein content during preparation. However, a faster preparation also led to a lumen fraction with less purity. Compared to the standard procedure the faster procedure showed an increased content of stromal proteins; after 2D-PAGE additional proteins were detected especially in the high molecular mass region.
**Thylakoid lumen proteome in dark versus light adapted plants**

*Arabidopsis thaliana* were grown in normal conditions in a growth-chamber at 100-150 photons m$^{-2}$ s$^{-1}$ with a growth cycle of 8/16 hours light/dark at 22/18°C. While leaves from dark adapted plants were harvested before the onset of light in the growth-chamber, leaves from light adapted plants were harvested 7.5 to 8 hours after the onset of light. To compare the protein content and expression of the two luminal preparations I used the DIGE technique (described in detail in the Key techniques section).

While 21 luminal protein spots were increased by almost 100% in the light sample, no spot was found to decrease. These data are in accordance with a manual analysis of publicly available transcriptomic data at BAR (Arabidopsis eFP Browser, bar.utoronto.ca). The 21 spots corresponded to 15 different proteins, several of them, as expected, have a function in photosynthesis (Paper II, Table 2A). One of the proteins with unknown function contains a putative bipartite transit peptide for transport to the lumen via the Tat-pathway and was detected in the thylakoid lumen for the first time. The up-regulation of this protein is small, but significant.

Co-expression analysis using publicly available microarray data further showed that the co-expression of lumen proteins was not limited to a light response but rather that the majority of the luminal protein genes are uniformly regulated on the transcriptional level (Paper II, Figure 3). This suggests a common function or participation in common biological pathways for the lumen proteins.

**Lumen proteins attachment to the thylakoid membrane**

Sodium bromide is a salt that releases proteins electrostatically attached to the thylakoid membrane and therefore increases the amount of proteins in the thylakoid lumen fraction. Before biochemically breaking the thylakoid membrane by yeda press, sodium bromide was added to the buffer and the resulting fraction of lumens proteins was compared to a preparation without sodium bromide. Due to the small changes the DeCyder™ program could not calculate any significant differences in relative protein amount, but still I was able to detect interesting changes. The newly proposed luminal protein At5g27390 as well as PsbP-1 and PsbQ-2 seem to be tightly attached to the membrane in dark adapted leafs; the relative amount of these proteins was increased when bromide was added to the yeda press buffer, figure 14.
Light regulation in plant

Light is perceived by photoreceptors and chloroplast redox signals. In addition the circadian clock regulates the plant to help it match the day and night cycle for optimal photosynthetic performance and thereby growing. The biological activity controlled by the circadian clock is cyclic with the highest activity at noon and the lowest at midnight. Because I took my samples in the morning before the light was turned on and in the afternoon before the light was turned off, the effect of circadian clock regulation on the proteins I found up-regulated, should be minimal. Instead the differences are more a result of photosynthetic acclimation and other means of diurnal regulation than the circadian clock.

Because many of the lumen proteins increase in abundance in the light, the total amount of proteins in the lumen will also be increased. Because the comparison between light and dark acclimated plants was done with the same total amount of protein, the observed expression changes probably are higher in reality. In order to really compare the differences in the lumen compartment in different conditions the number of chloroplasts should bee counted and then kept constant. This is not easy because the amount needs to be the same throughout the whole preparation and the chloroplasts should be intact. The weak chloroplast of Arabidopsis, as mentioned in paper I, are in that case not as suitable as Spinach chloroplasts.
Paper III

Isoforms and homologues in plants

Protein isoforms are proteins that have the same function and similar or identical sequence, but they are products of different genes. The isoforms originate from multiple gene copies, which means that there is more than one copy of the gene within the same genome. The large number of gene copies in Arabidopsis is explained by independent amplification of individual genes, and a mechanism involved could be unequal crossing over, and ancestral duplication of the entire genome. This is a problem in functional genomics as it is not enough to delete only one gene, instead you need to delete two or more, making functional genomics a problem. This is common in plants but has also been reported in humans. One protein in Arabidopsis could be due to as many as 25 different gene variations. The variations do not always result in differences in the protein though, because a difference in one nucleotide could still result in the same amino sequence. All variants are also not expressed because they are located downstream of a promotor which is closed temporarily or permanently. (The Arabidopsis Genome Initiative, 2000). Gene or protein sequences can be compared and homologues between plants can then indicate common origin. When we talk about protein homologues we refer to proteins showing a certain degree of similarity. This could be similarity in position, structure, function or characteristics.

Homologous isoforms of plastocyanin

Plastocyanin (PCp) is an important protein involved in the photosynthetic electron transport chain. Arabidopsis thaliana has two homologous plastocyanin isoforms (Kieselbach et al. 2000), encoded by the genes PETE1 and PETE2 in the nuclear genome. Several other angiosperms also have two isoforms of this soluble copper-binding protein, but this seems to be a result of a relatively recent evolutionary event in land plants, because only one plastocyanin isoform has been detected in cyanobacteria and algae. I got interested in the question of why plants like Arabidopsis have these two isoforms, PLAS1 (minor isoform) and PLAS2 (major isoform). Is it just an accident or does it have a functional role for the plant? Therefore I started to study the two isoforms of PCp. They have a very high identity and similarity, 82% and 92% and there is also a high identity between plastocyanins from other flowering plants, moss, algae and cyanobacteria. The difference between the two plastocyanin isoforms is only 18 of 99 amino acid residues. The residues Y83 and H87 (bold red in figure 15), are involved in electron transfer and are conserved both between the two plastocyanins in Arabidopsis and also in other species (Paper III, figure1). The same is true for D42, E43 and D44, (brown squares, figure 15) which comprise the acidic patch which is thought to be
involved in the reaction with PSI. The residues important for interaction with redox partners are G10, L12, G34, F35 and A90, (Blue dots, figure 15 and structure figure 16A). Here there are difference between the two PCp’s isoforms, PLAS2 has F35 instead of Y35 in PLAS1. This indicates a difference between the two isoforms in electron transfer. Electron transfer between PCp and P$_{700}^+$ was studied using Y35L mutant and wild type PCps from potato Haehnel et al. (1994). They found that electron transfer from PCp to P$_{700}^+$ was faster with the Y35L mutant, they also observed changes in the isoelectric point and binding constant to PSI, indicating changes in the protein fold. The amino acid differences between the two PCp isofoms from Arabidopsis also result in changes in isoelectric point, with a theoretical difference of 0.14 units and size difference of 6 Da. Almost all amino acid that differ between the two isoforms are located at the surface of the protein, (figure 16B).

Figure 15. Amino acid sequence of the two plastocyanin after cleavage of the bipartite transit peptide. Bold amino acid are amino acid that differ between the two isoforms. Red bold are amino acids which are involved in electron transfer. Blue dots are amino acids important for interactions with redox partners. The brown squares are residues thought to be involved in the reaction with PSI.

Figure 16. A. show amino acids in plastocyanin that are important for electron transfer, binding with cytochrome f and PSI. Figure B shows amino acids that differ between the two plastocyanin isoforms.
Movement of plastocyanin in the thylakoid lumen

Plastocyanin is quite mobile even in the thylakoid lumen, as it move between the Cyt $b_6/f$ complex and PSI. The Cyt $b_6/f$ complex is also mobile and in high light the complex moves to the grana region and so does also plastocyanin. In higher light more linear photo-phosphorylation occurs resulting in a higher amount of produced NADPH and carbon fixation. In the lumen the pH decreases due to the release of protons from the photosynthesis process. Because the proton transport taking place during photosynthesis is located in the grana, the luminal space in the grana region is probably more acidic than in the stromal lamellae. This leads to the surface charge density in the grana decreasing due to protonation of negative groups. This could be one explanation to why the PCp concentration increases in grana during illumination. The Cyt $b_6/f$ complex has a limited electron transport rate and therefore plastocyanin becomes oxidised in strong illumination. Reduced PCp molecules are probably bound to almost every PSI and the rest are mobile, but as soon PCp is oxidised it is released. This could be to protect PSI from photooxidative damage (paper III). Because of steric hindrance in the grana PSI are located in the stroma lamellae, end membranes in the margin of the grana. In low light there are more appressed grana, the PCp’s and the Cyt $b_6/f$ complex then move to the stroma lamellae and more cyclic electron flow occurs (Haehnel et al. 1989).

Cyt c6A and plastocyanin

Algae and cyanobacteria are thought to use either plastocyanin or Cyt $\alpha$ as electron donor to PSI, but in plants this protein was believed to been lost during the evolution of angiosperms, until a homologous Cyt $\alpha$ protein was found, Cyt $\alpha A$. However, it has been shown that Cyt $\alpha A$ could not donate electrons to PSI. Consequently, the only electron carrier between Cyt $b_6/f$ and PSI are the two plastocyanin isoforms (Weigel et al. 2003). Plastocyanin has also been proposed to interact with Cyt $\alpha A$ in thylakoid redox reactions, but we did not find any support for this, at least not a strong enough reaction for detection, (paper III).

Amount of plastocyanin

The amount of plastocyanin was compared in four plastocyanin mutants and wild type Arabidopsis thaliana (Columbia) by DIGE. Two of the mutants lack plastocyanin, one lacks the minor isoform, pete1, and the other the major isoform, pete2. Two of the mutants were overexpressors: one in the background of pete1 which over-expresses PLAS2 (PETE2), and the other in the background of pete2 which over-expresses PLAS1 (PETE1), table 1. The relative amount of plastocyanin protein was decreased by 90-95% in pete2, but this was still sufficient for growth (Pesaresi et al. 2008). In
Table 1. Name of the different Arabidopsis plastocyanin mutant and which protein that are expressed or not expressed in the plant.

<table>
<thead>
<tr>
<th>Name</th>
<th>Missing protein</th>
<th>Expressed protein</th>
<th>Alternative name</th>
</tr>
</thead>
<tbody>
<tr>
<td>pete1</td>
<td>∆PLAS1</td>
<td>PLAS2</td>
<td>pete1::En1</td>
</tr>
<tr>
<td>pete2</td>
<td>∆PLAS2</td>
<td>PLAS1</td>
<td>pete2::En1</td>
</tr>
<tr>
<td>PETE1</td>
<td>∆PLAS2</td>
<td>PLAS1 X 2</td>
<td>35S:PETE1</td>
</tr>
<tr>
<td>PETE2</td>
<td>∆PLAS1</td>
<td>PLAS2 X 2</td>
<td>35S:PETE2</td>
</tr>
</tbody>
</table>

paper III, pete2 showed a decrease in growth by ~30%. The pete1 mutant also shows a decrease in protein amount and growth. The corresponding over-expressors had a higher protein amount and increased growth compared to wild type and knockout mutants. The difference in growth between the mutants and wild type Arabidopsis seems to increase with growth time.

Photosynthesis performance in the mutants

The effective quantum yield of PSII (ΦII) was impaired and the rate of electron transport (ETR) under different actinic light intensities was only decreased in pete2. Reduction kinetic of P700+ also only show a decrease for pete2, which may be a result of the absence of pre-bound PCp. Oxidation kinetics of Cyt f also show a decrease for pete2 but in this case also for pete1 although to a lesser extent. Western analysis only showed reduction of major thylakoid multi-protein complexes in pete2. Although this is the mutant which differs most from the other mutants, it still has enough PCp to support photosynthetic electron flow. The two isoforms show a direct correlation between photosynthetic performance and the amount of PCp available. Therefore all data together implies that the two PCp isoforms, as predicted by their high sequence homology, are functionally equivalent.

The role of copper

Because PCp is a copper-binding protein and there was a large excess of PCp present in the thylakoid lumen under optimal growth conditions, the wild type and the four mutants were grown on solid agar medium with different concentration of copper. A copper content between 0.025 and 10 µM CuSO4 resulted in no major differences in growth, but in 25 µM CuSO4 pete2 shows growth reduction, implying that it has reached the threshold level for copper toxicity. As the over-expressors did not show any increased resistance to copper stress and pete mutants were more sensitive to oxidative stress, it is unlikely that PCp acts as a metallothioenin, mediating copper tolerance. Instead the large excess of PCp is thought to have a role as a copper sink.
The role of plastocyanin isoforms and amount

The plant has approximately 10% of PLAS1 and 90% PLAS2. The mutant pete2 therefore only has 10% PCp left. Even though it is negatively effected in all measurements, its limited photosynthetic performance and growth are only marginally evident under optimal growth conditions. The results of increased size and dry weight of leaves in over-expression mutants PETE1 and PETE2 could not be explained by increases in photosynthesis capacity. This could imply that the measurement for photosynthetic capacity has its limitations. If we disregard photosynthesis measurements, the large excess of PCp could be a way of ensuring an optimal lateral diffusion in the thylakoid lumen space between the grana and stroma lamellae, when the space is reduced in low light and different stress situations. PSI needs PCp for protection from photoxidative damage and therefore the plant needs more PCp than PSI. A surplus of PCp also seems to be for its role as copper sink in the plant.

Even if the pI differs and one of the amino acids important for redox partners differs (F35 and Y35), we did not observe that Arabidopsis two isoforms of PCp differ in photosynthetic performance. There could be two reasons for this: one is that the differences are not possible to measure with our technique to day, and the second could be due to conformation changes in the proteins, due to different amino acids, resulting in a similar electrostatic interaction in the thylakoid lumen.

Paper IV
Thylakoid lumen protein of 29kDa

TL29 is a luminal located protein with molecular mass of 29 kDa (Kieselbach et al. 2000), that is highly homologous to ascorbate peroxidases (APX). Based on this sequence homology it was renamed to APX4 and was suggested to protect the cell against reactive oxygen species (ROS) (Panchuk et al. 2002 and 2005 and Mittler et al. 2004). Usually APXs are heme and ascorbate binding enzymes that catalyse the reduction of H₂O₂ to water. Of the 9 APX enzymes in Arabidopsis three to four are predicted to be located in the chloroplast (Teixeira et al. 2006, Panchuk et al. 2005, Narendra et al. 2006). One of them, TL29, contains a twin-arginine motif prior to its hydrophobic region in the transit peptide and therefore was proposed to be transported via the Tat-pathway through the thylakoid membrane (Schubert et al. 2002). I was able to show that TL29 indeed is luminal located, it binds electrostatically to the grana fraction of the thylakoid membrane, but is easily washed of by 1M NaCl. Its location in regions, where mainly PSII is found suggested a
functional role within PSII. However comparing oxygen evolution and fluorescence in wild type Arabidopsis plants with a TL29 knock-out mutant did not indicate an involvement in the primary electron transport under normal growth conditions. Also the expression of other luminal proteins did not differ between wild type and knock-out mutant. 2D PAGE revealed four different isoforms of TL29 with approximal isoelectric points at pH 6, 6.5, 8.0 and 8.5, the most abundant spot being at pH 8.0, compared to the theoretical pI of TL29 is 7.87.

Although TL29 protein has high sequence similarity to ascorbate peroxidases, I could neither detect a distinct motive for heme binding nor amino acid residues typical for the active site of a peroxidise. Only one of the proposed key catalytic residues, L29, is present in TL29, but in an active enzyme the function of this residue is only to stabilise the binding of ascorbate (Sharp et al. 2003) More important is R38, which is missing in TL29. The two amino acids proposed to bind a second aromatic substance (Pipirou et al. 2007) are missing in TL29 as well. In tomato seven Apx genes were reported to code for true APX enzymes with all conserved amino acids that characterise APX catalytic sites and other essential sites. In contrary Arabidopsis has 6 true APX proteins, while APX4, APX6 and APX7 miss key amino acids (Najami et al. 2008).

In my studies I could not detect any ascorbate binding of TL29, neither using recombinant protein nor purified from Arabidopsis. A calorimetric method was used to measure ascorbate binding on recombinant enzymes. Recombinant APX1 (rAPX1) and PrxQ (rPrxQ), a luminal located peroxidase were used as controls and, as expected, high ascorbate affinity was measured for rAPX1, no affinity was detected for rPrxQ or rTL29. Removal of the N-terminal His6-tag of rTL29 abolished the week ascorbate affinity. Whole leaf extracts from Arabidopsis grown at low (50 µmol photons m⁻² s⁻¹), normal (120-150 µmol photons m⁻² s⁻¹) and high light (250-320 µmol photons m⁻² s⁻¹) showed higher ascorbate content after high light treatment, but no significant differences between wild type and the TL29 knock-out mutant. Although TL29 has high sequence similarity to APX enzymes I was able to show that TL29 is not an ascorbate peroxidase and therefore should be named TL29 and not APX4. Preliminary data obtained from a rTL29 crystal structure (Lundberg, Storm, Funk, Schröder, unpublished results) show that there are obvious structural differences between TL29 and other APXs. Instead, I received evidence that one of the cysteines in TL29 is redox-regulated. As only TL29 purified from thylakoids and not rTL29 showed redox-regulated function it might be that TL29 interacts with another protein or molecule at the chloroplast lumen side of PSII.
Summary

This thesis deals with the proteome of the chloroplast stroma and the thylakoid lumen and further analyses two luminal located proteins using functional genomics. I used the proteomic approach including typical tools like 2D-gel electrophoresis, difference in gel electrophoresis (DIGE) and also traditional biochemical analyses.

Paper I.

For the first time I present the 2D separation of the Spinach stromal proteome. An almost complete coverage was found for the enzymes of the Calvin cycle, and partial coverage of central metabolic pathways. I found that the cross-contamination of thylakoid lumen was less in Spinach compared to Arabidopsis, suggesting the chloroplast structure of Spinach being more robust. I found that the application of hierarchical clustering can increase the rate of identification for Spinach proteins to a similar extent as MS/MS and that clustering of mass spectra is a good starting point for detection of modified peptides with single amino acid substitution and post-translational modifications.

Paper II.

A fast procedure for the isolation of lumen fraction was developed, reducing the preparation time from 7 to 2 hours. Using this method luminal proteins isolated from plants grown at different light intensities were separated with DIGE. I found that in total 15 different lumen proteins were up-regulated in light, most of them are involved in the photosynthetic pathway. One of the up-regulated proteins (At5g27390) was detected for the first time in 2D-PAGE. A holistic co-expression analysis of all known lumen proteins revealed that they are uniformly transcriptional regulated. Lumen proteins therefore seem to have a common function or participate in common biological pathways.

Paper III

We show that the two isoforms of plastocyanin found in Arabidopsis have redundant functions in photosynthetic electron transport. Interestingly, we found that deletion of one of the plastocyanins lead to a 90% reduction of the total amount of plastocyanin, but only minor effects on the photosynthetic activity was seen in these plants. Plastocyanin might have an essential role in protecting higher plants from photo-oxidative damage. We further show that the two plastocyanins do not interact with cytochrome $c_6$. Instead we found that reduced plastocyanin levels increase the susceptibility to copper stress, suggesting that “overproduction” of plastocyanin could be explained in terms of a copper sink.

Paper IV

I show that the TL29 protein is located in the grana region of the thylakoid lumen, closely to the extrinsic oxygen evolving complex of Photosystem II. TL29 has homology to ascorbate peroxidises, but I could neither detect peroxidase activity, nor ascorbate or heme binding to the protein in vivo or in vitro. I did not find any function of TL29 in oxidative stress. TL29 is not an ascorbate peroxidase and should therefore not be entitled APX4. Interestingly, TL29 was found to be redox-regulated, implying interactions with another redox partner.
Perspective

Proteomics and protein identification – “High throughput” methods currently lead to a vast amount of data in a very short time. However, the field of functional genomics still is very time consuming. While protein identification has become easier with the technical development, finding functions for many of these proteins takes years of biochemical research. Sequence comparison of different proteins often can give suggestions for further biochemical analyses. However, general assumptions based on sequence comparisons might be misleading. The best example for misleading homology search is TL29. This protein has been renamed to APX4, its unusual expression has been analysed and described in several reviews. My biochemical studies now show that this protein has no peroxidase activity and does not even bind heme or ascorbate. Analysing every single protein will take time, it is estimated that today we only have biochemical data on roughly 10% of the 28,000 proteins encoded in the genome of Arabidopsis. Luckily better biochemical data on more proteins will improve the predictions as well.

SPECLUST - The use of SPECLUST and “peaks-in-common” will enable identification of proteins in organisms whose genome is not yet sequenced, or to search for post-translational modifications on the peptides of interest. A program should be developed to automate the manual inspection of peak lists before submitting for clustering and identification to simplify and speed up the process. The importance of post-translational modifications (except for phosphorylation) has been neglected for a long time and I think investigations on different post-modifications will help to get a deeper and better understanding of how organisms and cells are working and communicating. Specific questions that should be addressed are: I) What post-modifications exist in the chloroplasts? II) Can phosphorylation or glycosylation be found in the lumen? III) How is the communication between stroma and lumen occurring?

The thylakoid lumen – light as a regulator for photosynthetic proteins is more than obvious. However, why do all luminal proteins to seem be collectively regulated? Is there an additional regulation beside transcriptional expression? Are many luminal proteins redundant to cope with normal environmental changes? Are luminal proteins regulated due to complex formation and/or binding to the thylakoid membrane?

Also I would like to know why there are four isoforms of the TL29? What is the reason for this heterogeneity. My MALDI-TOF analyses showed that all isoforms have the same mass, and they are not phosphorylated. Maybe the crystal structure of TL29 will give an answer to this mystery.
Also the search for the possible redox-regulation partner of TL29 will be important and hopefully clarify the role of TL29 in the thylakoid lumen.

Plastocyanin has an essential role in photosynthesis and biomass production. The unexpected finding that plastocyanin is present at excess in plants, but still plastocyanin over-expressors increase plant size and dry weight, indicates that we still do not understand the complex function of this protein.

Other important questions concerning the thylakoid lumen that have to be solved in the future are: Are nucleotides present in the thylakoid lumen? Are there energy demanding activities or phosphorylations taking place in the lumen? There have been many thioredoxin targets detected in the lumen (Marchand et al. 2004 and 2006, Gopalan et al. 2004), where is the thioredoxin? Is there an APX in the lumen? There is also a large group of proteins in the thylakoid lumen with unknown function. To obtain biochemical data for all these will occupy PhD students for long time before we will be able to understand the dynamics of the thylakoid lumen and the role of this compartment in the photosynthetic process.
Acknowledgements

Thanks to all of you who have helped, encouraged and made this thesis possible. Especially I would like to express my sincere gratitude to:

Christiane who I got to know when I did the practical part in the Physiologic biochemistry course VT2002, when you had just moved to Umeå. Thank you for being positive even if I was not able to stress the cyanobacteria to produce Scps. Thank you for all help and support these years.

Mikael O. for telling me that there was a place as a PhD-student at the Department of Biochemistry.

My Supervisor Wolfgang for your supervising all these years and all the pep talks. It is a fact that I would not be here today without your tireless belief in my capability. Now I must stand on my own feet.

Wolfgangs group: Jose for all discussions and your positiveness. Good Luck with your PhD - it is your turn next time. Ulrika for letting me sleep at your place and for the plant, it grows but it is not as big as yours. Maria for collaboration on the TL29 project and kindness. Ruth for all your help and friendship. Liudmila for you kindness, friendship, helpfulness and encouragement. Good Luck with your PhD and life. Don’t be afraid of the Swedish bears, they eat blueberry and cowberry in the autumn.

Thomas for all discussions and all help and for letting me work in you lab.

Michael for collaboration, revising the language in my publications, for all interesting discussions, help, understanding, support and kindness.

Cecilia and Birger I really enjoyed the week I spent with you in Lund working with SPECLUST and I thank you for letting me stay in your house. Cecilia for your enthusiasm and fruitful collaboration.

Richard for introducing me to SPECLUST.

All members in Christianes group, both past and present; specially a grateful thanks to Maribel and Miguel for friendship and reading of my thesis before printing. Dmitry for help in understanding PAM measurement. Patrik for collaboration and for all help. Erik for discussion and I hope you will solve the TL29 structure.

Lars group, both past and present; Lars for co-supervision, keeping order in the corridor and teaching, you also taught me a lot before I started her. Barbara I witch I could walk with high heels like you, klick klock.

Anna-Märta for help with all administration questions, without you nothing will work.

Viola and Katarina for help with laboratory work and for lunch company, it is not the same now when you have left.
To all friends and colleagues both past and present at the Department of Biochemistry (even if it is not called that now) for having contributed to my thesis by being there at the coffee break, the lunch and in the corridors. I will never forget you.

Jakob and Anett for the days on Lanzarote and friendship

Personell at Kemiförrådet for being fun, positive and helpful.

Thanks to all my friends outside work, for reminding me that life is more than biochemistry, my thoughts are with you all.

Mamma och Pappa för att ni finns och alltid stöttar mig, Gunmari och Anki för ni är bäst, Vätternrundan nästa eller? Per, Mikael, Sara, Daniel, Isak och Lisa för att ni finns för mina systrar.

Måns thanks for believing in me, means more than you think and Vera, wish I had met you earlier.

My Family: Elin, Leila och Hugo for the fact that you are with me, I am very grateful for you letting me be your mother. It is a big honour. LOVE YOU ALL!!!

Svarte for you endurance, support and for just being you. LOVE YOU ☕️☕️☕️
References


Coleman, R. M. (1986). Wide awake at 3:00 a.m.: by choice or by change? New York.


Priesley, J. (1776). Experiments and Observations on Different Kinds of Air; The Discovery of Oxygen. Internet Modern History Sourcebook. 2nd. 2.


