The chloroplast lumen of Arabidopsis

Proteome map of the chloroplast lumen of *Arabidopsis thaliana*

Maria Schubert$^{1,2}$, Ulrika A. Petersson$^{2,3}$, Brian J. Haas$^4$, Christiane Funk$^3$, Wolfgang P. Schröder$^{2,5}$ and Thomas Kieselbach$^{1,6}$

$^1$Departments of Medical Nutrition and $^6$Biosciences, Karolinska Institute, Novum, SE-14186 Huddinge

$^2$Department of Natural Science, Södertörns University College, Bipontus, Box 4101, SE-14104 Huddinge

$^3$Department of Biochemistry, Arrhenius Laboratories for Natural Sciences, Stockholm University, SE-10691 Stockholm, Sweden

$^4$The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

$^5$Corresponding author, Tel + 46 8 58588587; Fax + 46 8 58588510
Email: wolfgang.schroder@sh.se

Running title: The chloroplast lumen of Arabidopsis
The chloroplast lumen of Arabidopsis

SUMMARY

The thylakoid membrane of the chloroplast is the centre of oxygenic photosynthesis. To better understand the function of the lumenal compartment within the thylakoid network, we have carried out a systematic characterization of the lumenal thylakoid proteins from the model organism Arabidopsis thaliana. Our data show that the thylakoid lumen has its own specific proteome, of which 36 proteins were identified. Besides a large group of peptidyl-prolyl cis-trans isomerases and proteases, a family of novel PsbP domain proteins was found. An analysis of the lumenal signal peptides showed that 19 of 36 lumenal precursors were marked by a twin-arginine motif for import via the Tat pathway. To compare the model organism Arabidopsis with another typical higher plant, we investigated the proteome from the thylakoid lumen of spinach and found that the lumenal proteins from both plants corresponded well. As a complement to our experimental investigation, we made a theoretical prediction of the lumenal proteins from the whole Arabidopsis genome and estimated that the thylakoid lumen of the chloroplast contains approximately 80 proteins.

Keywords

Arabidopsis genome/proteomics/bioinformatics/signal peptide/spinach
The chloroplast lumen of Arabidopsis

INTRODUCTION

The ability to perform oxygenic photosynthesis belongs to the distinguishing characteristics of higher plants, algae, and cyanobacteria. In higher plants the centre of the photosynthetic process is the thylakoid membrane of the chloroplast. Here, in a synergistic series of reactions, four protein complexes, the photosystems I and II, the cytochrome b\(_6\)f complex and the ATP-synthase produce NADPH and ATP that fuel the further synthesis of carbohydrates [1, 2].

A key feature in the energy conversion of photosynthesis is the link between the electron transfer from photosystem II to I via the cytochrome b\(_6\)f complex and the generation of a proton gradient over the thylakoid membrane. To balance the flow of electrical charges during the formation of the proton gradient, there is a busy traffic of chloride and calcium ions from the stroma into the lumen and of magnesium ions from the lumen into the stroma [3, 4, 5, 6]. This ion traffic plays a fundamental role for the proper function of photosynthesis. For a long time it was believed that accumulating protons and balancing the ion currents over the thylakoid membrane was the main function of the lumenal compartment. The ensemble of known lumenal proteins was small and consisted of the three extrinsic photosystem II-proteins PsbO, PsbP and PsbQ, and plastocyanin. This group was later joined by some new proteins such as violaxanthin de-epoxidase [7], polyphenol oxidase [8, 9], the extrinsic PSI-protein PsaN [10], and the carboxy-terminal processing protease for the D1-protein [11].

To achieve a more profound understanding of content and functions of the thylakoid lumen, we designed a method that enabled us to isolate a highly pure fraction of lumenal proteins from spinach thylakoids. For the first time, we showed that the lumen of the thylakoid
The chloroplast lumen of Arabidopsis membrane contained at least 20 proteins and that the protein concentration of this compartment was similar to that of the stroma [12]. Several new lumenal proteins could be characterised in more detail. The 17.4 kDa protein (TL17) had a remarkable new pentapeptide motif and led to the discovery of a whole family of unknown pentapeptide proteins in *Synechocystis sp. PCC 6803* [13]. In addition, a novel 16 kDa protein (TL16) was found to be routed into the thylakoid lumen by the Tat translocation pathway [14]. The lumenal immunophilin TL40 was suggested to participate in signal transduction over the thylakoid membrane [15], and *Hcf136* was identified as a lumenal assembly factor for photosystem II [16]. Recently, the 29 kDa peroxidase homologue TL29 and a novel plastocyanin (PLAT) were added to the list of new lumenal proteins [17] and several not further characterized proteins from the thylakoids of the chloroplast from pea [18].

The completion of the *Arabidopsis thaliana* genome sequencing project by the end of 2000 [19] started a new era in plant research. To apply the knowledge of the Arabidopsis genome to an investigation into lumenal proteins, we designed a method to isolate the lumenal proteins from Arabidopsis chloroplasts and studied them by proteomics. In this study, we performed the first systematic characterisation of the chloroplast lumen of Arabidopsis and compared its proteins with those from the chloroplast lumen of spinach. We found 36 lumenal proteins in Arabidopsis, of which 22 could be identified in spinach, too. By comparing the experimentally identified lumen proteins of Arabidopsis with a theoretical prediction of a lumenal proteome in this organism, we estimated that the chloroplast lumen of Arabidopsis comprises approximately 80 proteins.
EXPERIMENTAL PROCEDURES

Growth conditions and lumen preparation

Plants were cultivated hydroponically according to [20]. Arabidopsis thaliana ecotype Columbia was grown for 13 weeks with eight hours light per day and a light intensity of 100 µmole photons m\(^{-2}\) s\(^{-1}\), and spinach (Spinacia oleracea) was raised for five weeks with 10 h light per day. The lumen fraction from spinach chloroplasts was isolated according to [12]. The lumenal content from Arabidopsis chloroplasts was isolated in the same way with the following minor changes. To avoid protein degradation, 1 mM of EDTA was added to the last washing buffer and 50 µg/ml of Pefablock was added to the thylakoids just prior to the Yedapress treatment. The lumen fraction was concentrated using Centriprep YM-3 concentrators (Millipore), and protein quantification was carried out according to [21] using bovine serum albumin as a standard.

2D electrophoresis

The lumenal proteins were separated by isoelectric focusing in the first and by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension. The samples for analytical 2D gels contained 100 µg of protein and those for Western blots for micro sequencing 200 µg protein. The lumenal proteins were solubilized in 5 M urea, 2 M thiourea, 4% (w/v) CHAPS, 50 mM DTT and 0.8 % (v/v) carrier ampholytes (IPG-buffer pH 3-10NL or 4-7L, (Amersham Pharmacia Biotech) and applied during rehydration to a non-linear IPG-stripped pH 3-10 or a to linear IPG-strip pH 4-7. The strips were allowed to rehydrate over night at 20°C and then transferred to IPGphor cup-loading strip-holders and covered with paraffin
The chloroplast lumen of Arabidopsis

oil. The proteins were focused for 10 minutes at 300 V followed by a three-hours gradient from 300 V to 3500 V and a 30-minutes gradient from 3500 V to 8000 V. The isoelectric focusing was then completed at a constant voltage of 8000 V until 60,000 Vhr was reached. Subsequently, the strips were equilibrated first for 15 min in 50 mM Tris-HCl (pH 6.8), 6M urea, 30% v/v glycerol, 2% w/v SDS and 1% (w/v) DTT and then for 10 min in the same buffer without DTT but with 2.5% (w/v) iodoacetamide and a trace of bromphenol blue. In the second dimension, SDS-PAGE according to [22] was carried out in a gradient slab gel (T=9-16%) using a Protean II XL system from BioRad. Before the polymerisation of the gel, 5 mM of sodium thiosulphate was added to the monomer solution to decrease the background staining with silver. Proteins were detected by silver staining according to [23]. The 2D gels were scanned using an image scanner and evaluated with the Image Master 2D software (both from Amersham Pharmacia Biotech). The apparent masses of the proteins that were detected on the 2D gels were determined manually or with the Image master 2D software using identified proteins of known masses as a reference.

MALDI-TOF mass spectrometry and micro sequencing

MALDI-TOF analysis of in-gel-digested proteins was carried out with a Reflex III mass spectrometer from Bruker. The in-gel-digests were performed using sequencing-grade modified trypsin (Promega) and analysed as described [24, 25]. Database searches were done with the MS BioTools software from Bruker using the Mascot search engine (available on-line at the address: www.matrixscience.com). If a protein could not unambiguously be identified by a fingerprint spectrum, its identity was confirmed by a post source decay analysis of single peptides [26]. Amino-terminal micro sequencing was carried out with a Procise sequencer
The chloroplast lumen of Arabidopsis from Applied Biosystems. Proteins were sequenced from polyvinylidene difluoride membrane following resolution by 2D electrophoresis essentially as described [27].

**Bioinformatics**

The individual analysis of proteins sequences by similarity searches [28, 29], pattern and profile searches [30, 31], alignments [32], and hydrophobicity plots [33] was carried out using the ExPaSy tools (available on-line at www.expasy.ch). The prediction of chloroplast targeted proteins encoded within the Arabidopsis thaliana genome was performed by subjecting the latest version of the proteome, currently consisting of 25,657 proteins, to an analysis using the program TargetP via the web interface [34]. The N-terminal portion of each protein sequence, not exceeding 140 residues in length, was analyzed by the plant version of TargetP, and all chloroplast predicted proteins (rank 1 to 5) were used. The prediction of signal peptides and the peptide cleavage products was performed using the portable version of SignalP-2.0 [35, 36]. The programs TargetP and SignalP are available online at the address www.cbs.dtu.dk/services, and the Arabidopsis proteome is accessible at the address: ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/SEQUENCES/ATH1.pep

**RESULTS**

**Isolation of the thylakoid lumen, reproducibility of the 2D electrophoresis and protein identification**

While Arabidopsis is an excellent model organism for molecular biological studies, biochemical work has been difficult due to the small leaf size and the low amount of material.
The chloroplast lumen of Arabidopsis introduction of a hydroponical culture made it possible to overcome this drawback and to grow plants with considerably larger leaves that provided sufficient amounts of plant material in a high quality well suited for biochemical preparations. An essential element in this technique was the restriction of light to eight hours per day, which enabled to grow the plants for 13 weeks without flowering. Using Arabidopsis plants cultivated in this way, we were able to purify the luminal fraction in the same high quality as it was obtained in the original method for the isolation of thylakoid lumen from spinach chloroplasts [12]. A typical lumen preparation from Arabidopsis started with 100-200 g of dry weight leaf material and yielded in approximately 2 mg protein per 100 g of leaves.

An important prerequisite for a concise mapping of the luminal Arabidopsis proteins was a reproducible two dimensional (2D) electrophoresis system that was capable of resolving the major part of the luminal proteins. To meet these requirements we used a combination of non-linear pH gradients (pH 3-10) and a polyacrylamide gradient gel that allowed us to detect proteins with isoelectric points between 4 and 9 and masses between 200 and 9 kDa. Samples from 14 different lumen preparations were analysed in more than 20 experiments. The 2D maps of the luminal Arabidopsis proteins showed an excellent reproducibility, and a representative experiment is shown in figure 1A. From the complete set of experiments, thirteen 2D gels were selected for a detailed image analysis. The total number of protein spots that were detected on the 2D gels was between 400 and 700, and 277 of those were present on all 13 gels.
The chloroplast lumen of Arabidopsis

The protein pattern of the 2D gel in figure 1A shows clearly that most proteins were detected in the acidic region of the pH gradient, while the group of basic proteins was relatively small. It should be noticed, too, that there is a distinct gap between the proteins in the acidic and the basic region of the pH gradient, where only few proteins could be detected. Since 80% of the lumenal proteins were found in the acidic range of the pH gradient, we analysed these proteins in more detail using a 2D electrophoresis system with a linear pH gradient from 4 to 7. A typical 2D map of the lumenal proteins with isoelectric points in this pH range is shown in figure 1B. An evaluation of the images of 10 independent experiments showed that approximately 200 protein spots were present on each 2D gel of the pH range 4 to 7.

Having established a reproducible 2D map of the chloroplast lumen from Arabidopsis, we systematically analysed the lumenal proteins by both MALDI-TOF mass spectrometry and amino-terminal micro sequencing. Using this combination of methods we were able to determine the amino-termini of the mature proteins and to correct errors in the gene models of several proteins. This approach could be successfully applied to 90 protein spots, but the others of the 277 spots that were detected in all 2D gels were too weak for an analysis by both mass spectrometry and micro sequencing. In total, 49 proteins were detected, and each protein was identified in at least two independent experiments except for the three putative fibrillins with the apparent masses at 25.5 kDa, 25.3 kDa and 24.7 kDa. Although these proteins were analysed only once, the identification was specific in each single case, and the corresponding protein spots were detected in all 2D gels analysed. Hence, there is no doubt about that these proteins were correctly identified.
The chloroplast lumen of Arabidopsis

The 2D gel in figure 1A shows that 40 of the 49 identified proteins were detected in the acidic range of the pH gradient while 9 proteins were found in the basic region. The analysis of the proteins by mass spectrometry and micro sequencing showed that the major part of the proteins from the lumenal fraction was intact. There were only few degradation products found. As figure 1B shows there are two fragments of PsbP1 that were detected in all 2D gels analysed. In addition, we also detected a degradation product for PsaN, but this fragment occurred less frequent and did not appear in a distinct pattern as those from PsbP1.

The protein content of the chloroplast lumen of Arabidopsis

One of the principal objectives of this study was to identify a representative group of proteins from the chloroplast lumen of Arabidopsis and to find indications for their possible functions. Hence, it was important to confirm that the proteins that were identified in the lumenal fraction from Arabidopsis chloroplasts were truly lumen located. The purification method alone was no proof for a lumenal location. As figure 1B shows, we detected among the lumenal proteins a putative glyoxylase and the Cyp4 cyclophilin that are putative stroma proteins but were present in similar amounts as known lumenal proteins such as TL29 and Hcf136.

To resolve whether the proteins that were identified in the lumenal fraction were resident in the chloroplast lumen or not, we performed a concise analysis of their transit peptides. All proteins that are targeted to the chloroplast lumen are synthesized in the cytosol as precursors and cross the chloroplast envelope and the thylakoid membrane in a two-stage import via a stromal intermediate. Accordingly, the signal peptides of lumenal precursors
The chloroplast lumen of Arabidopsis comprise two parts, of which one is designed for the transit through the envelope and the second one for the import into the thylakoid lumen. Once the precursor of a lumenal protein has reached the chloroplast stroma, the envelope transit region is cleaved off, and the intermediate is routed into the thylakoid lumen by either the Sec pathway or the ΔpH-dependent twin-arginine translocation (Tat) pathway. While the Sec machinery is used by proteins that tolerate partial unfolding during the import, proteins that need to be imported in a folded state cross the thylakoid membrane via the Tat complex [37].

The bipartite transit peptides for the import into the thylakoid lumen are a distinguishing feature of the lumenal proteins. They are marked by a hydrophilic, serine and threonine rich region for the transit through the envelope and a thylakoid targeting region with a typical hydrophobic core close to the processing site. In addition, the transit peptides of proteins that are routed by the Tat complex reveal a distinctive twin-arginine motif in the beginning of the hydrophobic core region and a highly hydrophobic residue two or three positions after the twin-arginine motif. By contrast, the transit peptides of proteins that are translocated by the Sec-machinery do not have a twin-arginine motif but a single lysine residue next to the amino-terminal end of the hydrophobic core region [37, 38].

The transit peptides of all proteins that were identified in the lumenal fraction of Arabidopsis chloroplasts were comprehensively analysed whether they revealed the features of bipartite signal peptides or not. Of the 49 proteins that were identified in the lumenal fraction, 35 had, indeed, a bipartite transit peptide. Figure 2 shows the thylakoid targeting regions of these transit peptides aligned with the program Clustal W. The alignment shows that all transit
The chloroplast lumen of Arabidopsis peptides possess a hydrophobic core; 19 signal peptides have a twin-arginine motif that marks them for translocation by the Tat pathway, and 16 have a lysine residue close to the hydrophobic core which is a characteristic of signal peptides routed by the Sec-machinery. Only the D1 processing protease appears to be an exception from this rule and has an arginine instead of lysine residue next to the amino-terminal end of the hydrophobic core.

The only protein for which it was hard to decide, if the transit peptide really contained a targeting signal for the thylakoid lumen was the 20 kDa protein (Q9LXX5). The prepeptide of this protein had a rather indistinct hydrophobic core region and an arginine triplet instead of a conventional twin-arginine motif. To resolve this apparent conflict, we searched the est-databases for homologues and examined their transit peptides. The soybean clone Gm-c1032-2020 and the tomato clone cLET42E20 encoded for full-length homologues of the 20 kDa-protein and had both a typical bipartite transit peptides with a plain twin-arginine motif (not shown). That ensured that the prepeptide of the 20 kDa protein contained only an unusual variant of the twin-arginine motif, and, indeed, was a true bipartite transit peptide.

In case of the 19 kDa protein (P82658), no Arabidopsis gene was available that could be used for an examination of the signal peptide. It only could be mapped to the tentative consensus sequence TC115875 in the Arabidopsis Gene Index, which can not be used to assess its subcellular location. Hence, we searched the est databases and found several cDNAs from other plants that encoded the complete precursor of a homologue of the 19 kDa protein. Two representative examples were the tomato cDNA cTOF22B19 and the soybean cDNA Gm-c1013-3374. An analysis of the signal peptides of these homologues demonstrated clearly that
The chloroplast lumen of Arabidopsis

they had all features of a bipartite transit peptide (not shown), which indicates that the 19 kDa protein from Arabidopsis is a lumen located protein, too. The other 14 proteins that were identified in the lumenal fraction from Arabidopsis came either from the chloroplast stroma such as the small subunit of ribulose-bisphosphate carboxylase or the thylakoid membrane such as the Ftsh-protease or the alpha- and beta subunits of the ATP synthase. As for the four putative fibrillins and the 26 kDa protein the subcellular location could not be predicted reliably, and, hence, it was left undecided.

As table 1 shows, the entire range of the lumenal proteins of Arabidopsis that were identified in this work covers both well established classical proteins such as violaxanthin de-epoxidase and the extrinsic subunits of photosystem II, as well as novel proteins such as the large group of cyclophilins and FKBP-type peptidyl-prolyl cis-trans isomerases. In addition, there are many new proteins for which yet no function is known as, for instance, the PsbP domain proteins.

The silver-stained 2D gels in figures 1A and 1B show clearly that the major proteins in the lumenal fraction from Arabidopsis were the isoforms of the extrinsic subunits of photosystem II and of plastocyanin. As shown in table 1, the Arabidopsis genome encodes for two similar forms of each of these proteins, and we identified all of them except for the protein PSP2 that despite strong efforts could not be found. Figure 1B shows plainly that the protein pattern reveals a marked isoelectric heterogeneity for these proteins. The extrinsic protein PSP1, for instance, was found in as many as four spots excluding those two that contained degraded forms of PSP1. However, we have not yet investigated whether this heterogeneity represents
The chloroplast lumen of Arabidopsis post-translational modifications of the single protein or just an artificial effect of the 2D electrophoresis system. Remarkably, we also found a group of six proteins with apparent masses between 15.9 kDa and 35.8 kDa that have a PsbP domain that is related to the extrinsic PsbP protein of photosystem II.

Two further lumenal proteins with functions for photosystem II were found. The xanthophyll cycle enzyme violaxanthin de-epoxidase (VDE) participates in the protection of photosystem II from excess light [39], and Hcf136 (H136) is essential for photosystem II assembly [16]. The positions of these proteins are marked in figures 1A and 1B at the apparent molecular masses of 44.2 kDa and 34.8 kDa. As with the extrinsic subunits of photosystem II both Hcf136 and violaxanthin de-epoxidase revealed isoelectric heterogeneity and were detected in three different spots (figure 1B). A distinct isoelectric heterogeneity was also observed for the putative ascorbate peroxidase TL29, the pentapeptide protein TL17, and for the 40 kDa protein Q9SSA5 that is a homologue of the TL40 immunophilin of spinach.

Proteases are important regulatory proteins and their presence in the thylakoid lumen was postulated for a long time. DegP and the D1-processing protease were the first proteases that were found to be lumen located [40, 11]. In this work, three novel lumenal proteases were identified that include a carboxy-terminal proteinase D1-like protein with an apparent mass of 49 kDa, a DegP-like 36 kDa protein, and the 27 kDa HhoA protease. As figure 1B shows, isoelectric heterogeneity was observed for this group of proteins, too, except for the HhoA protease.
The chloroplast lumen of Arabidopsis

Immunophilins play an important role in protein folding, and comprise the largest group of the novel luminal proteins shown in table I. They consist of two groups. There are three putative cyclophilin-type peptidyl-prolyl cis-trans isomerases (PPIase) that were identified at the apparent molecular masses of 40 kDa, 38 kDa, and 18.5 kDa, and five putative FKBP-type PPIases with apparent molecular masses between 14.7 kDa and 18.5 kDa. It is noteworthy that five of these putative isomerases were basic proteins. Figure 1 reveals that only the 38 kDa and the 40 kDa protein that belong to the putative cyclophilin-type PPIases, and the 17.5 kDa FKBP-type PPIase isolog were detected in the acidic range of the 2D gels.

As for the other 7 luminal proteins that were identified, no functions are known yet. Two pentapeptide repeat proteins were identified, of which TL17 was the first higher plant protein that possessed this motif [13]. The 11.6 kDa protein that was found in this study is a novel member of this group. In addition, we identified five further proteins with apparent molecular masses between 15 kDa and 19 kDa that do not reveal any known motifs, patterns or domains.

The protein content of the chloroplast lumen of spinach

As Arabidopsis is a prominent model organism, we wanted to know, if its luminal protein ensemble was representative for other higher plants. The organism of choice for this study was spinach, because the preparation of the luminal content of the spinach chloroplast was thoroughly characterized and known to produce a highly pure fraction of luminal proteins [12]. To obtain a typical 2D map of the luminal spinach proteome, a set of nine independent 2D experiments was carried out, and one representative 2D gel is shown in figure 3. The
The chloroplast lumen of Arabidopsis protein pattern revealed more than 500 spots, of which approximately 250 were detected in all nine 2D experiments. A comparison of figures 3 and 1A shows that the 2D patterns of the lumenal proteins from spinach and Arabidopsis chloroplasts are very similar. As with Arabidopsis the major part of lumenal proteins from spinach was detected in the acidic range of the pH gradient, and there is a similar gap between the acidic and the basic proteins as it was found for the lumenal proteins from Arabidopsis. Although the 2D pattern of the lumen proteins from spinach resembles much the one from Arabidopsis, it is obvious that it reveals considerably less isoelectric heterogeneity.

As no complete genome is available for spinach, protein identification was mainly carried out by micro sequencing, which restricted the analysis of the lumenal proteome from spinach to the more abundant proteins. Nevertheless, 25 lumenal proteins from the spinach chloroplast were identified, and Table II shows that they correspond well to the lumenal proteins from Arabidopsis. All groups of proteins that were found in the thylakoid lumen of Arabidopsis were also identified in the lumenal compartment of the spinach chloroplast. Some differences were observed, however. Among the lumenal proteins from spinach, we could not identify any isoforms of plastocyanin and the extrinsic photosystem II subunits PsbO and PsbQ. In addition, we did not find any Arabidopsis homologues to the 25.3 kDa protein and polyphenol oxidase. While we could not resolve experimentally, if the 25.3 kDa protein was really absent in Arabidopsis, we could analyse the lumenal content of the Arabidopsis chloroplast for the presence of polyphenol oxidase activity. In two preparations of lumen fraction from Arabidopsis no polyphenol oxidase activity was detectable (data not shown), and, as with garden pea, this enzyme seems not to be present in Arabidopsis. However, for 22
The chloroplast lumen of Arabidopsis

of the lumenal proteins from the spinach chloroplast a homologue in Arabidopsis was identified. In addition, there are 12 homologous proteins from garden pea [18] that can be related to both the lumenal proteins from spinach and Arabidopsis. All these data indicate that the lumenal compartments of these three plants essentially comprise the same proteins.

**Prediction of the lumenal chloroplast proteins of Arabidopsis**

As a complement to our experimental characterization of the thylakoid lumen from Arabidopsis, we made a theoretical prediction of the lumenal proteins from the whole Arabidopsis proteome to estimate the entire number of proteins in the thylakoid lumen. To detect the lumenal proteins within the Arabidopsis proteome, we used the bipartite transit peptides as a marker because they contain the information that is necessary to target these proteins to the chloroplast lumen and, hence, are specific for their subcellular location. The prediction was carried out in three steps. First, we screened the latest version of the complete fasta database of Arabidopsis proteins with the program TargetP for chloroplast located proteins. From a total number of 25,657 Arabidopsis proteins, 3765 were predicted to have its subcellular location in the chloroplast. In the second stage, we screened the entire TargetP predicted chloroplast proteome for potential signal peptides using the program SignalP 2.0 and obtained 514 signal peptides that were at least 30 residues in length. Among those, 358 had a length between 50 and 120 residues and, thus, were candidates for transit peptides of lumenal proteins. Finally, the 358 preselected sequences were examined manually for typical features of bipartite transit peptides such as a hydrophilic, serine and threonine rich amino-terminal region and a hydrophobic domain close to the processing site. In addition, the complete precursors were scanned for the presence of known pattern and profiles to exclude proteins that could not reside in the thylakoid lumen such as ion channels or chlorophyll
The chloroplast lumen of Arabidopsis binding antennae proteins. To avoid false positives and to keep the number of missed luminal proteins to a minimum the manual analysis of the SignalP selected sequences was carried out two times, and conflicts between the two prediction cycles were examined again in each single case. From the 358 sequences that were individually analysed, 303 were excluded, and 55 were predicted to belong to putative luminal proteins. The result from this evaluation is summarized in Table III and a summary of the single steps of the prediction is provided in Table IV.

To evaluate the prediction, we used the 35 experimentally identified luminal proteins, for which full-length sequences were available. The 19 kDa protein, for which no gene is known yet, had to be excluded. The TargetP predicted chloroplast proteome contained 34 of the identified luminal proteins. Violaxanthin de-epoxidase was missed by TargetP, which probably was due to its unusually long transit peptide. From the 34 identified luminal protein that were present in the predicted chloroplast proteome, 25 were predicted to be lumen located. The main reason why 9 proteins escaped prediction was the preselection of putative luminal proteins with SignalP that failed to recognize these proteins. Remarkably, all their precursors but one had a signal peptide with a twin-arginine motif, which indicates that we particularly underestimated proteins of the Tat pathway. In summary, the prediction found 25 of the 35 experimentally identified luminal proteins that were present in the Arabidopsis proteome, which was a rate of 71%. If we assume that these proteins are a representative part of the luminal proteome, we can extrapolate from the number of 55 predicted luminal proteins and estimate that the thylakoid lumen comprises approximately 80 proteins.
DISCUSSION

The investigation into the chloroplast lumen from Arabidopsis and from spinach that was carried out in this work showed clearly that the lumenal proteins from both plants correspond well and the lumenal proteome from Arabidopsis can serve as a model for other higher plants. To better understand the proteins from the chloroplast lumen, we compared our data with the annotated Arabidopsis proteome at the European Bioinformatics Institute (EBI) that is available on-line at the address www.ebi.ac.uk/proteome. This analysis showed that the proteins that were found in the chloroplast lumen do not contain any members of the 10 biggest protein families in Arabidopsis such as eukaryotic protein kinases or proteins of the F-box domain family. By contrast, the lumenal proteins that were identified belong to families that consist of a rather low number of proteins, of which, however, a relatively large part is located in the thylakoid lumen. A good example for the specific pattern of the lumenal proteome are the lumenal immunophilins. From a total number of 22 FKBP-type peptidyl-prolyl cis-trans isomerase (PPIases) that are annotated in the Arabidopsis proteome at the EBI, five were detected in the chloroplast lumen. In addition, 4 further FKBP-type PPIases were found among the predicted lumenal proteins (table III). The group of lumenal proteases has a specific composition, too. It includes all known members of the family of tail specific proteases that are the D1-processing protease, the processing protease D1-like protein, and the putative 45 kDa protein Q9SVY2. The latter was not identified in this work but found among the predicted proteins (table III). In addition, there are 3 lumenal proteases that belong to the trypsin family of the serine proteases, of which 15 members are annotated in the Arabidopsis proteome. We did not find any member of one of the big families of the
The chloroplast lumen of Arabidopsis

subtilisin-type serine proteases with 53 annotated proteins, and the sumo-specific proteases with 62 annotated proteins. As with the pentapeptide repeat proteins that are a big family in bacteria and cyanobacteria [41], only 4 proteins with this motif are annotated in the Arabidopsis proteome. Two of those, the 17 kDa protein TL17 and the 11.6 kDa protein T116, were identified in the thylakoid lumen, and a third one, the putative 24 kDa protein Q9FWX1, was predicted to be lumen located (Table III). Although the functions of these proteins are unknown, they probably play a role that is specifically needed in the thylakoid lumen. Furthermore, the photosystem II assembly factor Hcf136 and the enzyme violaxanthin de-epoxidase, contribute to the specific composition of the luminal proteome. The sequence of Hcf136 possesses a BNR repeat motif that, for instance, is found in many bacterial and eukaryotic glycosyl hydrolases but only in 2 Arabidopsis proteins. Violaxanthin de-epoxidase has a lipocalin motif, which is very rare among the proteins of Arabidopsis but, for example, is present in many eukaryotic fatty acid binding proteins. In addition, there is a novel family of PsbP domain proteins of which no members in other subcellular compartments are known. Besides the known PsbP1 protein of photosystem II, 6 further members of this family were identified, and 2 more were found among the predicted luminal proteins. The large number of these proteins indicates that they fulfil an important function in the thylakoid lumen. In summary, all these data show that the chloroplast lumen of Arabidopsis not only contains a large number of proteins but also has its own specific proteome.

The identified and the predicted luminal proteins of Arabidopsis in tables I and III correspond well, and this remarkable agreement indicates that both the experimental and the in silico approach have covered a representative part of the luminal proteins. An important result of
The chloroplast lumen of Arabidopsis

the *in silico* approach was the discovery of 2 putative PsbQ related proteins that provide a complement to the family of PsbP domain proteins. The 24.8 kDa precursor Q9SGH4 and the 22.2 kDa precursor Q9XI73 possess the Prodom domain PD007524 that also is a typical feature of the extrinsic photosystem II subunits PsbQ1 and PsbQ2 from Arabidopsis. This domain is only shared by the currently known PsbQ proteins from higher plants, which indicates that it is a characteristic of those proteins. The extrinsic proteins PsbO, PsbP, and PsbQ play an important role in the established model of photosystem II from higher plants [42, 2, 43]. It is generally accepted that these proteins participate in the regulation of oxygen evolution and are present in photosystem II in a stoichiometric ratio of 1:1:1. While it has not been entirely clarified whether the photosystem II complex has one or two copies of each extrinsic protein [2], recent crystal structure data indicated that there is one copy of the PsbO protein per photosystem II [43]. The discovery of novel PsbP domain proteins in this study suggests that the lumenal surface of photosystem II might have a more complex composition than previously believed. Classical reconstitution experiments with thylakoids from spinach showed that photosystem II does not require the PsbP protein to produce oxygen. However, it has a considerably higher oxygen evolving activity, if the PsbP protein is bound to the thylakoid membrane [44]. Consistent results were obtained from a study of two types of photosystem II complexes from spinach, of which one contained only the extrinsic PsbO protein and the other one both the PsbO and PsbP protein [45]. The novel PsbP domain proteins of the thylakoid lumen might fulfil similar functions and provide for photosystem II a tool to modulate its oxygen evolving activity.
The chloroplast lumen of Arabidopsis

While the extrinsic PsbO protein is found in the photosystem II of both higher plants and cyanobacteria, the PsbP and PsbQ protein are only known in higher plants, and it is believed that the cyanobacterial proteins PsbU and PsbV fulfil the function of the higher plant proteins PsbP and PsbQ [46]. However, the hypothetical 20.7 kDa protein P73952 from the cyanobacterium *Synechocystis sp*. PCC 6803 has a distinct PsbP domain, and it is a homologue to the PsbP-like T215 protein from Arabidopsis. Currently, the function of the 20.7 kDa protein is unknown, but deletion mutants in *Synechocystis sp*. PCC6803 have been performed and studies are under progress to find out whether this protein plays a role in photosynthesis or not (Funk and Schröder, unpublished data).

Besides the PsbP domain proteins, immunophilins from the families of the cyclophilin- and FKBP-type PPIases were the biggest group of proteins that was found in the chloroplast lumen of Arabidopsis. As a general feature, immunophilins have the ability to catalyse the cis-trans isomerization of proline-imidic peptide bonds and to accelerate protein folding [47, 48]. Remarkably, the chloroplast lumen contains an unusually big group of FKBP-type PPIases. An analysis of the signal peptides of all TargetP-predicted chloroplast immunophilins from the annotated Arabidopsis proteome at the EBI indicated that the FKBP-type PPIases of the chloroplast only occur in the thylakoid lumen (data not shown). That implies that these proteins fulfil rather specific than general functions that are particularly needed in the thylakoid lumen. In contrast to their mammalian relatives, plant immunophilins are a relatively recent discovery, and few of them have been functionally characterized. The only lumenal immunophilin that was studied in detail is the cyclophilin TL40 from spinach that regulates the dephosphorylation of thylakoid membrane proteins by binding to a PP2A-like
The chloroplast lumen of Arabidopsis

phosphatase [49, 50]. The functions of the other lumenal immunophilins are currently unknown, but these proteins could participate in processes such as activation or inhibition of other thylakoid proteins, and protein folding and assembly. A putative target group for the activity of lumenal immunophilins are, for instance, the extrinsic proteins PsbO, PsbP and PsbQ. These proteins show the unusual behaviour that they not only bind to the thylakoid membrane but also exist in the lumen in an unassembled state, in which they are long-lived and assembly competent [51, 52]. As unassembled proteins usually are rapidly degraded, it could be hypothesized that lumenal immunophilins participate in the protection of these unassembled extrinsic proteins and in their proper assembly to photosystem II.

The novel lumenal proteins from Arabidopsis do not only show that the chloroplast lumen has a characteristic proteome but also illustrate the impact of the ΔpH-dependent Tat-pathway for the protein traffic into this compartment. As figure 2 demonstrates, 19 of 35 lumenal protein precursors have a signal peptide with a twin-arginine motif and are marked for translocation by the Tat-complex. That indicates that more than half of the lumenal proteins from Arabidopsis might be routed across the thylakoid membrane via the Tat-pathway. In combining the experimentally identified 35 thylakoid lumen proteins, for which genes are known, with the 55 predicted ones, there are a total of 66 Arabidopsis proteins that are presumed to occupy the thylakoid lumen. The genes for these proteins appear to be near evenly distributed among all 5 Arabidopsis chromosomes: 20 within chr. 1, 9 within chr2, 14 within chr. 3, 10 within chr. 4, and 13 within chr. 5. In addition, they appear to be evenly distributed within each chromosome as well, and only 4 genes are within 100 kb of another gene encoding a presumed thylakoid targeted protein. It should be noted, too, that none of the
The chloroplast lumen of Arabidopsis predicted or identified lumenal proteins possessed any known ATP binding sites, and, thus, this work does not confirm previous reports on the presence of Hsp70 and related proteins in the thylakoid lumen [53, 18]. In earlier studies, we analysed the thylakoid lumen of spinach for the presence of ATPase activity [12, 54], but the detectable activity was so low that it was not specific for the lumenal proteins. That does not exclude that ATP or other nucleotides are used in the chloroplast lumen, but so far the evidence for this has not been provided, and the source of energy for the lumenal proteins is still unknown.

The annotation of protein coding genes within the Arabidopsis genome is heterogeneous, and the accuracy of each annotated gene is based on the available evidence that supports it. That includes homology to protein and nucleotide sequences, and in the extreme case, it is limited to computationally predicted genes. Most gene annotations are based on a combination of methods. The importance of experimentally determined sequence data in providing more accurate annotations is greatly exemplified by this work. When we searched the Arabidopsis genome for the experimentally identified lumenal proteins, we found 6 genes which required gene structure annotation refinements that altered the annotated protein coding sequence of the gene. TIGR’s Arabidopsis genome reannotation efforts are greatly benefiting from such genomic and proteomic studies. The identification of these genes as proteins of the chloroplast lumen also provides experimentally determined cellular localization data which can be represented within the context of Gene Ontology assignments [55]. The lumenal 19 kDa protein, which could not be mapped to the Arabidopsis genome, demonstrates the need for continued sequencing efforts to further complete the Arabidopsis genome. Such efforts are
The chloroplast lumen of Arabidopsis currently in progress at TIGR, and a high priority has been established to identify the sequence containing the gene for this 19 kDa protein.

In summary, this study has shown that the chloroplast lumen not only fulfils a function for the generation of the pH gradient that fuels ATP synthesis but also has its own specific proteome. In the chloroplast lumen from Arabidopsis, 36 proteins were identified, and the entire luminal proteome of Arabidopsis was estimated to comprise approximately 80 proteins. That suggests that the narrow luminal space of the thylakoid membrane is densely packed with proteins. The discovery of the novel PsbP domain proteins and immunophilins in the chloroplast lumen of Arabidopsis indicates that luminal proteins play an important role for the regulation of photosynthesis.

**ACKNOWLEDGEMENTS**

We would like to thank Professor Jan-Åke Gustafsson (Karolinska Institute, Huddinge) for his support and Dr Fredrik Nilsson (Astra Zeneca AB, Göteborg) for his advice and discussions concerning mass spectrometry analysis. This project was supported by grants of Swedish Natural Science Research Council and Södertörns Högskola. Brian J. Haas was supported by Cooperative Agreement DBI-9813586 from the National Science Foundation. Parts of this work were carried out at the Protein Analysis Unit at the Centre of Structural Biochemistry at the Karolinska Institute.
The chloroplast lumen of Arabidopsis

REFERENCES


The chloroplast lumen of Arabidopsis


The chloroplast lumen of Arabidopsis


The chloroplast lumen of Arabidopsis


The chloroplast lumen of Arabidopsis

LEGENDS TO FIGURES

Fig. 1A. Silver-stained 2D-gel of 100 µg of soluble luminal proteins from the chloroplast of Arabidopsis. The proteins were resolved by SDS-electrophoresis in a 9-16% polyacrylamide gradient gel subsequent to isoelectric focusing in a non-linear immobilized pH gradient from pH 3 to 10.

Fig. 1B. Silver-stained 2D-gel of 100 µg of soluble luminal proteins from the chloroplast of Arabidopsis. The proteins were resolved by SDS-electrophoresis in a 9-16% polyacrylamide gradient gel subsequent to isoelectric focusing in a non-linear immobilized pH gradient from pH 4 to 7.

Fig. 2. Transit peptides of luminal proteins from Arabidopsis thaliana. The alignment shows the hydrophobic core region (underlined) and the processing site of the bipartite transit peptides of the luminal proteins from Table 1. Transit peptides of proteins that are routed by the tat complex possess a distinctive twin arginine motif. By contrast, transit peptides of proteins that are targeted by the Sec-pathway have a lysine residue close to the hydrophobic core region.

Fig. 3. Silver-stained 2D-gel of 100 µg of soluble luminal proteins from the chloroplast of spinach. The proteins were resolved by SDS-electrophoresis in a 9-16% polyacrylamide gradient gel subsequent to isoelectric focusing in a non-linear immobilized pH gradient from pH 3 to 10.
### TABLES

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Gene locus</th>
<th>Accession</th>
<th>Identification</th>
<th>Mass [kDa] exper./theor.</th>
<th>Experimental N-terminal sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Xanthophyll cycle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violaxanthin de-epoxidase</td>
<td>AT1g08550, Q39249</td>
<td>1, 3</td>
<td>44.2 / 39.8</td>
<td>VDALKTCACL LGKCRIELAK CIANPACAXN</td>
<td></td>
</tr>
<tr>
<td><strong>Photosystem II assembly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H136_ARATH</td>
<td>AT5g23120, O82660</td>
<td>1, 3</td>
<td>Not determined / 35.8</td>
<td>DEQLSEWERV FLPID</td>
<td></td>
</tr>
<tr>
<td><strong>Photosystem II subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSO1_ARATH</td>
<td>AT5g66570, P23321</td>
<td>1, 3</td>
<td>Not determined / 26.6</td>
<td>EGAPKRLT</td>
<td></td>
</tr>
<tr>
<td>PSO2_ARATH</td>
<td>AT3g50820, Q9S841</td>
<td>1, 3</td>
<td>24.7 / 26.6</td>
<td>EGAPKRLT</td>
<td></td>
</tr>
<tr>
<td>PSP1_ARATH</td>
<td>AT1g06680, Q42029</td>
<td>1, 3</td>
<td>Not determined / 20.2</td>
<td>AYGEAANVFG KPKTN</td>
<td></td>
</tr>
<tr>
<td>PSP2_ARATH</td>
<td>AT2G30790, O49344</td>
<td>Not detected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSQ1_ARATH</td>
<td>AT4g21280, Q9XFT3</td>
<td>1, 3</td>
<td>Not determined / 16.3</td>
<td>DAISIKVGPP PAPS</td>
<td></td>
</tr>
<tr>
<td>PSQ2_ARATH</td>
<td>AT4g05180, Q41932</td>
<td>1, 3</td>
<td>15.3 / 16.3</td>
<td>EAPIKVGGGP PLPS</td>
<td></td>
</tr>
<tr>
<td><strong>Proteins with a PsbP domain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.8 kDa protein</td>
<td>AT5g11450, P82715</td>
<td>1, 3</td>
<td>24.3 / 25.6</td>
<td>EEDVKMSGEE LKMGTMVDDI</td>
<td></td>
</tr>
<tr>
<td>TL30_ARATH</td>
<td>AT1g77090, O49292</td>
<td>1, 3</td>
<td>19.4 / 22.2</td>
<td>VVKQGGLAGR VPGLSEPDE</td>
<td></td>
</tr>
<tr>
<td>T215_ARATH</td>
<td>AT4g15510, Q23403</td>
<td>1, 3</td>
<td>21.5 / 21.3</td>
<td>STPVFREYID TFDGYSFKYP</td>
<td></td>
</tr>
<tr>
<td>TL26_ARATH</td>
<td>AT3g55330, P82538</td>
<td>1, 3</td>
<td>16.3 / 17.8</td>
<td>AESKKGFLAV SDNKDAYAFL YPFGWQEVEVI EGQDKV</td>
<td></td>
</tr>
<tr>
<td>20 kDa protein</td>
<td>AT3g56650, Q9LXX5</td>
<td>1, 3</td>
<td>20.0 / 21.5</td>
<td>REVEVGSYL PLSDDPXFL</td>
<td></td>
</tr>
<tr>
<td>15.9 kDa protein</td>
<td>AT1g76450, Q9S720</td>
<td>1, 3</td>
<td>15.9 / 15.8</td>
<td>ETNASEAFRV YTDETNKFEI SIPQ</td>
<td></td>
</tr>
<tr>
<td><strong>Plastocyanins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAS_ARATH</td>
<td>AT1g76100, P11490</td>
<td>3</td>
<td>Not determined / 10.5</td>
<td>MEVLLGSD</td>
<td></td>
</tr>
<tr>
<td>PLAT_ARATH</td>
<td>AT1g20340, P42699</td>
<td>3</td>
<td>16.2 / 10.5</td>
<td>IEVLLGGGDG SLAFIPNDFS IAKGEKIVF</td>
<td></td>
</tr>
<tr>
<td><strong>Putative peroxidase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL29_ARATH</td>
<td>AT4g09010, P82281</td>
<td>1, 3</td>
<td>23.2 / 29.3</td>
<td>ADLNQRRQRS EFQSIXILL STTIKAKPEL</td>
<td></td>
</tr>
<tr>
<td><strong>Proteases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tail specific proteases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxy-terminal proteinase D1-like protein</td>
<td>AT5g46390, Q9FL23</td>
<td>1, 3</td>
<td>49.0 / 45.8</td>
<td>ATNDPYLS</td>
<td></td>
</tr>
<tr>
<td>D1 processing protease</td>
<td>AT4g17740, Q23614</td>
<td>1, 3</td>
<td>39.6 / 41.9</td>
<td>LTEENLLFXE A</td>
<td></td>
</tr>
</tbody>
</table>
The chloroplast lumen of Arabidopsis

<table>
<thead>
<tr>
<th>Serine proteases, trypsin family</th>
<th>AT5g39830, Q9LU10</th>
<th>1, 2, 3</th>
<th>31.3 / 37.5</th>
<th>LGDPsvATVE DVSPtvFPAG PLF</th>
</tr>
</thead>
<tbody>
<tr>
<td>DegP protease</td>
<td>AT3g27925 a)</td>
<td>1, 3</td>
<td>31.7 / 35.2</td>
<td>FVVSTPpKQLQ TDELA</td>
</tr>
<tr>
<td>HhoA protease</td>
<td>AT4g18370 a)</td>
<td>1, 3</td>
<td>27.4 / 23.5</td>
<td>LEQFKEXEEX L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Putative immunophilins</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclophilin-type PPIases</td>
<td>AT3g01480, Q9SSA5</td>
<td>1, 2, 3</td>
<td>36.1 / 38.2</td>
<td>VLISGPPIKD PEALLRYALPID</td>
</tr>
<tr>
<td>FKBP-type PPIases</td>
<td>AT5g13410, Q9LYR5</td>
<td>1, 3</td>
<td>18.0 / 18.7</td>
<td>SQFADMPALK GKDYGKTKMK YPDY</td>
</tr>
<tr>
<td></td>
<td>AT1g20810, Q9LM71</td>
<td>1</td>
<td>17.8 / 17.9</td>
<td>RERRSRKVIP</td>
</tr>
<tr>
<td></td>
<td>AT2g43560 a)</td>
<td>1, 3</td>
<td>17.5 / 15.7</td>
<td>AGLPPEDKPR LCEAXCXXL</td>
</tr>
<tr>
<td></td>
<td>AT3g60370, Q9M222</td>
<td>1, 3</td>
<td>16.9 / 16.4</td>
<td>KTKSKSPYDE RRLLEQN</td>
</tr>
<tr>
<td>Pentapeptide proteins</td>
<td>AT5g53490, P81760</td>
<td>1, 3</td>
<td>15.4 / 17.4</td>
<td>ANQRLPLPLST EPDR</td>
</tr>
<tr>
<td>Proteins with unknown function</td>
<td>AT2g444920, O22160</td>
<td>1, 3</td>
<td>14.7 / 11.5</td>
<td>FKGGGPYGQG VTRGQDLSGK</td>
</tr>
<tr>
<td>Proteins with undecided location</td>
<td>AT5g52970, Q9LV5</td>
<td>1, 3</td>
<td>15.0 / 11.5</td>
<td>KVGVNKPELL PKEFTSVIDV</td>
</tr>
<tr>
<td>Putative fibrillins</td>
<td>AT4g04020, O81439</td>
<td>1, 3</td>
<td>26.7 / 28.8</td>
<td>ATDIDDE</td>
</tr>
<tr>
<td></td>
<td>AT4g22240, O49629</td>
<td>1</td>
<td>25.5 / not</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td>AT3g23400, Q9LW57</td>
<td>1</td>
<td>25.3 / not</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td>AT3g58010, Q9M2P7</td>
<td>1</td>
<td>24.7 / not</td>
<td>Not</td>
</tr>
</tbody>
</table>

| Putative immunophilins           | AT5g35100, O65220 | 1, 3   | 25.5 / 27.9 | TVTPPPAKP PSPDITIDR      |

The chloroplast lumen of Arabidopsis contains various serine proteases, particularly those belonging to the trypsin family. The DegP and HhoA proteases are also present. Putative immunophilins, such as cyclophilin-type PPIases, are found in the chloroplast lumen. FKBP-type PPIases, which play a role in protein folding, are also identified. Pentapeptide proteins are another class of proteins found in the chloroplast lumen. Proteins with unknown function and proteins with an undecided location are also present, including putative fibrillins and immunophilins.
The chloroplast lumen of Arabidopsis

<table>
<thead>
<tr>
<th>Protein Description</th>
<th>Accession</th>
<th>MALDI-TOF-MS</th>
<th>PSD</th>
<th>Micro sequencing</th>
<th>Swiss-Prot/TrEMBL</th>
<th>Tigr Arabidopsis db</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATPase alpha chain</td>
<td>AtpA, P56757</td>
<td>Not determined / 55.3</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT2g30950, O80860</td>
<td>1, 3</td>
<td>FGQSXAXF (residues 209-216)</td>
</tr>
<tr>
<td>ATPase beta chain</td>
<td>AtpB, P19366</td>
<td>Not determined / 53.9</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g50250, Q9SX43</td>
<td>1</td>
<td>Not determined</td>
</tr>
<tr>
<td>Putative ftsh</td>
<td>Not determined</td>
<td>51.8 / not determined</td>
<td>49.9 / not determined</td>
<td>Not determined</td>
<td>AT2g30950, O80860</td>
<td>1, 3</td>
<td>Not determined</td>
</tr>
<tr>
<td>Chloroplast protease</td>
<td>FTSH_ARATH</td>
<td>27.9 / not determined</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g50250, Q9SX43</td>
<td>1</td>
<td>FGQSXAXF (residues 209-216)</td>
</tr>
<tr>
<td>Ferredoxin-NADP+</td>
<td>AT5g66190, Q9FKW6</td>
<td>25.0 / 33.2</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g57280, Q9FYG5</td>
<td>1, 3</td>
<td>Not determined</td>
</tr>
<tr>
<td>Putative glyoxylase</td>
<td>AT1g57280, Q9FYG5</td>
<td>31.3 / 38.1</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g50250, Q9SX43</td>
<td>1, 3</td>
<td>FGQSXAXF (residues 209-216)</td>
</tr>
<tr>
<td>Fructose bis-</td>
<td>AT2g21330, Not available</td>
<td>16.8 / 19.8</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT3g62030, P34791</td>
<td>1, 3</td>
<td>Not determined</td>
</tr>
<tr>
<td>phosphate aldolase</td>
<td>CYP4_ARATH</td>
<td>14.5 / not determined</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g57280, Q9FYG5</td>
<td>1</td>
<td>FGQSXAXF (residues 209-216)</td>
</tr>
<tr>
<td>RBS small subunit</td>
<td>AT1g57280, Q9FYG5</td>
<td>31.3 / 38.1</td>
<td>Not determined</td>
<td>Not determined</td>
<td>AT1g57280, Q9FYG5</td>
<td>1, 3</td>
<td>FGQSXAXF (residues 209-216)</td>
</tr>
</tbody>
</table>

1 MALDI-TOF-MS
2 PSD
3 Micro sequencing
4 Swiss-Prot/TrEMBL
5 Tigr Arabidopsis db
a) corrected in this work
<table>
<thead>
<tr>
<th>Protein name</th>
<th>Accession number</th>
<th>Identification</th>
<th>Exp. Mass [kDa]</th>
<th>Experimental N-terminal sequence</th>
<th>Homologue in Arabidopsis</th>
<th>Homologue in garden pea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Xanthophyll cycle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violaxanthin de-epoxidase</td>
<td>Q9SM43</td>
<td>1, 3</td>
<td>43.1</td>
<td>VDALKTCTXL</td>
<td>Violaxanthin de-epox. (Q39249)</td>
<td></td>
</tr>
<tr>
<td><strong>Photosystem II assembly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative HCF136</td>
<td>P82714</td>
<td>2, 3</td>
<td>34.6</td>
<td>EDSLSDWERV YLPIDPGVVL</td>
<td>H136_ARATH (O82660)</td>
<td>37.3 kDa protein (P82342)</td>
</tr>
<tr>
<td><strong>Photosystem II subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSBO_SPIOL</td>
<td>P12359</td>
<td>1, 3</td>
<td>Not determined</td>
<td>EGGKRLTY</td>
<td>PSO1_ARATH (P23321)</td>
<td>PSBO_PEA (P14226)</td>
</tr>
<tr>
<td>PSBP_SPIOL</td>
<td>P12302</td>
<td>1, 3</td>
<td>Not determined</td>
<td>AYGESA</td>
<td>PSP1_ARATH (Q42029)</td>
<td>PSBP_PEA (P16059)</td>
</tr>
<tr>
<td>PSBQ_SPIOL</td>
<td>P12301</td>
<td>1, 3</td>
<td>Not determined</td>
<td>EARPIVGPPE PPLSG</td>
<td>PSQ1_ARATH (Q9XFT3)</td>
<td>PSBQ_PEA (P19589)</td>
</tr>
<tr>
<td><strong>Putative PsbP domain proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 kDa protein</td>
<td>P82796</td>
<td>3</td>
<td>22.0</td>
<td>EQSAGFREYI DFFDGYSLTY</td>
<td>T215_ARATH (O23403)</td>
<td>24 kDa protein (P82337)</td>
</tr>
<tr>
<td>TL1X_SPIOL</td>
<td>P82537</td>
<td>3</td>
<td>17.0</td>
<td>AESKKGFLPV IDKKDGYTFL YPGFGQEVSI</td>
<td>TL26_ARATH (P82538)</td>
<td>16.9 kDa protein (P82340)</td>
</tr>
<tr>
<td>20 kDa protein</td>
<td>P83090</td>
<td>3</td>
<td>20.0</td>
<td>RDVDVGSFLP KSPSDPSMVL</td>
<td>20 kDa protein (Q9LXX5)</td>
<td>24.1 kDa protein (P82329)</td>
</tr>
<tr>
<td><strong>Plastocyanins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLAS_SPIOL</td>
<td>P42699</td>
<td>1, 2, 3</td>
<td>Not determined</td>
<td>VEVLLGG</td>
<td>PLAT_ARATH (P42699)</td>
<td>PLAS_PEA (P16002)</td>
</tr>
<tr>
<td><strong>Putative peroxidase</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL29_SPIOL</td>
<td>P81833</td>
<td>3</td>
<td>23.1</td>
<td>ADLIQRRQXS</td>
<td>TL29_ARATH (P82281)</td>
<td>28.5 kDa protein (P82338)</td>
</tr>
</tbody>
</table>
The chloroplast lumen of Arabidopsis

<table>
<thead>
<tr>
<th>Putative proteases</th>
<th>Putative DEGP protease</th>
<th>36.3</th>
<th>FVVSTPXKLQ</th>
<th>DEGP_ARATH (O22609)</th>
<th>46 kDa protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immunophilins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TL40_SPIOL</td>
<td>O49939</td>
<td>1, 3</td>
<td>Not deter-</td>
<td>VLISGPP</td>
<td>40 kDa protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative immunophilins</td>
<td>33.6 kDa protein</td>
<td>3</td>
<td>33.6</td>
<td>VLYSPDTKVP R</td>
<td>38 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P83092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TL18_SPIOL</td>
<td>3</td>
<td>17.7</td>
<td>SAEETPLQSK VTNKVVFDIS</td>
<td>18 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P82536</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.8 kDa protein</td>
<td>3</td>
<td>17.8</td>
<td>AGLPPEEKPK LCAACE</td>
<td>17.5 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P83061</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative penta-peptide proteins</td>
<td>TL17_SPIOL</td>
<td>3</td>
<td>13.5</td>
<td>ANQRLPPLSN DPDCRAFV GNT</td>
<td>18.3 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P81778</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11 kDa protein</td>
<td>3</td>
<td>11.2</td>
<td>FKGGGQPYGGG VTRGQDLGSK DF</td>
<td>11.6 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P82657</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.3 kDa protein</td>
<td>3</td>
<td>18.3</td>
<td>DLNKFEAEMR GEFGIXSA</td>
<td>Putative 30.1 kDa protein (Q9FWX1)</td>
</tr>
<tr>
<td>Proteins with unknown function</td>
<td>Polyphenol oxidase</td>
<td>3</td>
<td>Not deter-</td>
<td>APLILPDVEKCT LSDALWDG</td>
<td>Not found</td>
</tr>
<tr>
<td></td>
<td>P43310</td>
<td></td>
<td>mined</td>
<td>AIANAPLLDT TITDRVFFD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.3 kDa protein</td>
<td>3</td>
<td>25.3</td>
<td></td>
<td>Not present in this plant</td>
</tr>
<tr>
<td></td>
<td>P83089</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.4 kDa protein</td>
<td>3</td>
<td>18.4</td>
<td>DSPTPNYNI YYGTAASAXN</td>
<td>19 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P82799</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TL16_SPIOL</td>
<td>3</td>
<td>13.3</td>
<td>APLEDEDDLE LLEKVKRDRK KRLERQGAIN</td>
<td>16 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P81834</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.8 kDa protein</td>
<td>3</td>
<td>13.8</td>
<td>LDEFRVYSSD ANKYKISIPQ D</td>
<td>15.9 kDa protein</td>
</tr>
<tr>
<td></td>
<td>P82681</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Downloaded from http://www.jbc.org/ by guest on January 28, 2018
The chloroplast lumen of Arabidopsis

<table>
<thead>
<tr>
<th>Protein</th>
<th>Accession</th>
<th>MW (kDa)</th>
<th>Amino acid sequence</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.7 kDa protein</td>
<td>P82682</td>
<td>14.7</td>
<td>KTGVNKPELL PKEETTVIDV</td>
<td></td>
</tr>
<tr>
<td>9 kDa protein</td>
<td>P82671</td>
<td>9.1</td>
<td>GFLSGSTGIE XIPGPQL</td>
<td>Putative 11.9 kDa protein (AT2g03420)</td>
</tr>
</tbody>
</table>

**Stromal and other proteins**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Accession</th>
<th>MW (kDa)</th>
<th>Amino acid sequence</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>FENR_SPIOL</td>
<td>P00455</td>
<td>31.2</td>
<td>QIASDVEA</td>
<td></td>
</tr>
</tbody>
</table>

1 MALDI-TOF-MS
2 PSD
3 Micro sequencing
4 Swiss-Prot/TrEMBL
5 Tigr Arabidopsis db

a) Gene corrected in this work
b) According to [18]
<table>
<thead>
<tr>
<th>Gene locus</th>
<th>SignalP-NN prediction</th>
<th>Predicted processing site</th>
<th>Protein name and accession</th>
<th>Putative import pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photosystem II assembly</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5g23120</td>
<td>78</td>
<td>78_79 ARA_DE</td>
<td>H136_ARATH (O82660)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>Photosystem II subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3g56820</td>
<td>86</td>
<td>84_86 AGA_EG</td>
<td>PSO2_ARATH (Q9S841)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT5g66570</td>
<td>85</td>
<td>85_86 ASA_EG</td>
<td>PSO1_ARATH (P23321)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT4g05180</td>
<td>82</td>
<td>82_83 VFA_EA</td>
<td>PSO2_ARATH (Q41932)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g06680</td>
<td>77</td>
<td>77_78 ADA_AY</td>
<td>PSP1_ARATH (Q42029)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT4g21280</td>
<td>75</td>
<td>75_76 VLA_DA</td>
<td>PSO1_ARATH (Q9XFT3)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT2g30790</td>
<td>70</td>
<td>70_71 AIG_SK</td>
<td>PSP2_ARATH (O49344)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>Putative PsbP domain proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4g15510</td>
<td>104</td>
<td>104_105 AFA_ST</td>
<td>T215_ARATH (O23403)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT5g11450</td>
<td>95</td>
<td>95_96 THA_EE</td>
<td>35.8 kDa protein (P82715)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g76450</td>
<td>80</td>
<td>80_81 AFA_ET</td>
<td>15.9 kDa protein (Q9S720)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT3g55330</td>
<td>74</td>
<td>74_75 SFA_AE</td>
<td>TL26_ARATH (P82538)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT2g39470</td>
<td>73</td>
<td>73_74 LLA_EE</td>
<td>P18.6 (O80634)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>Putative PsbQ domain proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3g01440</td>
<td>67</td>
<td>74_75 SLA_QD</td>
<td>P24.8 (Q9SGH4)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g14150</td>
<td>65</td>
<td>65_66 ALA_QE</td>
<td>P22.2 (Q9XI73)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>Plastocyanins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1g76100</td>
<td>72</td>
<td>72_73 AMA_ME</td>
<td>PLAS_ARATH (P11490)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g20340</td>
<td>68</td>
<td>68_69 AMA_IE</td>
<td>PLAT_ARATH (P42699)</td>
<td>Sec</td>
</tr>
<tr>
<td><strong>Photosystem I subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5g64040</td>
<td>86</td>
<td>86_78 ANA_GV</td>
<td>PSAN_ARATH (P49107)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>Proteases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tail specific proteases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4g17740</td>
<td>119</td>
<td>119_120 VTT_DS</td>
<td>D1 processing protease (Q23614; Q9ZP02)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT3g57680</td>
<td>99</td>
<td>99_100 ALA_ES</td>
<td>P45.4 (Q9S720)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT5g46390</td>
<td>68</td>
<td>68_69 VAA_TN</td>
<td>Carboxy-terminal protease D1 like protein (Q9FL23)</td>
<td>Sec</td>
</tr>
<tr>
<td><strong>Serine proteases, trypsin family</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3g27925</td>
<td>102</td>
<td>105_106 ASA_FV</td>
<td>DEGP_ARATH (O22609)</td>
<td>Sec</td>
</tr>
<tr>
<td><strong>Immunophilins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cyclophilin-type peptidyl-prolyl cis-trans isomerasers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3g01520</td>
<td>103</td>
<td>114_115 ALA_VL</td>
<td>38 kDa protein (P82869)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT3g01480</td>
<td>82</td>
<td>91_92 DVS_VL</td>
<td>40 kDa protein (Q9SSA5)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g74070</td>
<td>74</td>
<td>74_75 AQA_DT</td>
<td>P22.2 (Q9C9C7)</td>
<td>Tat</td>
</tr>
<tr>
<td><strong>FKBP-type peptidyl-prolyl cis-trans isomerasers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1g18170</td>
<td>94</td>
<td>94_95 VIS_EQ</td>
<td>P16.5 (Q9LDY5)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT3g10060</td>
<td>82</td>
<td>82_83 AEA_VS</td>
<td>P17.8 (Q9SR70)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT4g19830</td>
<td>78</td>
<td>78_79 ALA_DF</td>
<td>P16.7 (Q9SR70)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT2g43560</td>
<td>73</td>
<td>76_77 AYA_AG</td>
<td>17.5 kDa protein (O22870)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT4g39710</td>
<td>73</td>
<td>73_74 ADA_TR</td>
<td>P12 (O65658)</td>
<td>Sec</td>
</tr>
</tbody>
</table>
The chloroplast lumen of Arabidopsis

<table>
<thead>
<tr>
<th>Accession</th>
<th>MW (kDa)</th>
<th>Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1g20810</td>
<td>17.8</td>
<td>71_72 SEA_RE 16.9 protein (Q9LM71)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT3g60370</td>
<td>53</td>
<td>53_54 SSS_AK 16.9 protein (Q9M222)</td>
<td>Tat</td>
</tr>
</tbody>
</table>

### Pentapeptide proteins

<table>
<thead>
<tr>
<th>Accession</th>
<th>MW (kDa)</th>
<th>Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2g44920</td>
<td>81</td>
<td>81_82 ALA_FG 11.6 kDa protein (Q22160)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g12250</td>
<td>79</td>
<td>83_84 AMA_EL P19.2 (Q9FWX1)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT5g53490</td>
<td>77</td>
<td>77_78 VIA_AN TL17_ARATH (P81760)</td>
<td>Sec</td>
</tr>
</tbody>
</table>

### Unknown function

<table>
<thead>
<tr>
<th>Accession</th>
<th>MW (kDa)</th>
<th>Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT3g44020</td>
<td>103</td>
<td>70_71 SSA_FD P17.1 (Q9LXV9)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT2g40400</td>
<td>93</td>
<td>93_94 ASA_ET P71 (Q9SIY5)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT2g37400</td>
<td>91</td>
<td>91_92 AIA_AP P28.3 (Q9ZUS6)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT3g56140</td>
<td>90</td>
<td>90_91 SLA_SE P73.5 (Q9LYM7)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g33780</td>
<td>88</td>
<td>88_89 GDA_SQ P28.3 (Q9LQ30)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g54780</td>
<td>84</td>
<td>84_85 ALA_SE 18.3 kDa protein (Q9ZVL6)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT3g09490</td>
<td>84</td>
<td>81_82 FSA_SF P38.4 (Q9SF54)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT5g02590</td>
<td>84</td>
<td>84_85 VTA_AT P36.5 (Q9LZ43)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT2g26340</td>
<td>80</td>
<td>80_81 AMA_GG P19.1 (Q48702)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g51400</td>
<td>76</td>
<td>76_77 AMA_DD P3.3 (Q9SYE2)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT5g52970</td>
<td>75</td>
<td>75_76 ADA_KV 15.0 kDa protein (Q9LVV5)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g79450</td>
<td>74</td>
<td>71_72 LFA_SQ P37.9 (Q9SAK3)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g51350</td>
<td>73</td>
<td>73_74 ADA_DD P65.8 (Q9SY7D)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT2g23670</td>
<td>71</td>
<td>71_72 SMA_EN P10 (Q64835)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g03600</td>
<td>68</td>
<td>68_69 VSA_AE P11.7 (Q9LR64)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT1g62140</td>
<td>66</td>
<td>64_65 EMA_AA P40.1 (Q04580)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT4g24930</td>
<td>63</td>
<td>63_64 ALA_IP 17.9 kDa protein (Q9SWS3)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g14590</td>
<td>62</td>
<td>61_62 YRA_AD P38 (Q9MA24)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT5g46560</td>
<td>59</td>
<td>59_60 AVA_FT P36.4 (not available)</td>
<td>Sec</td>
</tr>
<tr>
<td>AT1g21500</td>
<td>59</td>
<td>59_60 ALA_AK P13.2 (Q9LPK9)</td>
<td>Tat</td>
</tr>
<tr>
<td>AT3g08550</td>
<td>57</td>
<td>57_58 GLA_DP P53.8 (Q9C9Z9)</td>
<td>Sec</td>
</tr>
</tbody>
</table>

a) identified in this work; b) conflict between the sequence in the Tigr and in the Swiss-Prot/Trembl database
### Table IV Summary of the prediction of lumenal chloroplast proteins of *Arabidopsis thaliana*

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis proteome</td>
<td>25,657</td>
</tr>
<tr>
<td>TargetP predicted chloroplast proteome</td>
<td>3765</td>
</tr>
<tr>
<td>Predicted lumenal proteins</td>
<td>55</td>
</tr>
<tr>
<td>Experimentally identified lumenal proteins</td>
<td>36</td>
</tr>
<tr>
<td>Predicted and experimentally identified lumenal proteins (overlap excluded)</td>
<td>66</td>
</tr>
<tr>
<td>Estimated number of lumenal proteins</td>
<td>80</td>
</tr>
</tbody>
</table>

- Estimated number of lumenal proteins: 80 proteins, which is 0.3% of the Arabidopsis proteome and 2% of the putative chloroplast proteome.
Thylakoid lumen of Arabidopsis thaliana
Thylakoid lumen of Arabidopsis thaliana
Fig 2.

**TAT pathway**

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>sp</td>
<td>O82660</td>
<td>H136_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>Q42029</td>
<td>PSP1_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>Q9XT3</td>
<td>PSQ1_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>Q41932</td>
<td>PSQ2_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>P82715</td>
<td>35.8_kDa_protein</td>
</tr>
<tr>
<td>sp</td>
<td>O49292</td>
<td>TL30_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>P82538</td>
<td>TL26_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>O23403</td>
<td>T215_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>P82281</td>
<td>TL29_ARATH</td>
</tr>
<tr>
<td>tr</td>
<td>Q9U10</td>
<td>36_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9LY5</td>
<td>18_kDa_protein</td>
</tr>
<tr>
<td>sp</td>
<td>O22780</td>
<td>17.5_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9M222</td>
<td>16.9_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9SCY5</td>
<td>14.7_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9S20</td>
<td>15.9_kDa_protein</td>
</tr>
<tr>
<td><strong>:</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sec pathway**

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>tr</td>
<td>Q7249</td>
<td>Violaxanthin_de</td>
</tr>
<tr>
<td>sp</td>
<td>P23321</td>
<td>PS01_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>Q9841</td>
<td>PS02_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>P11490</td>
<td>PLAS_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>P42699</td>
<td>PLAT_ARATH</td>
</tr>
<tr>
<td>sp</td>
<td>O22609</td>
<td>DEGP_ARATH</td>
</tr>
<tr>
<td>tr</td>
<td>Q92P02</td>
<td>D1_processing</td>
</tr>
<tr>
<td>tr</td>
<td>P82869</td>
<td>38_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9AS6</td>
<td>18.5_kDa_protein</td>
</tr>
<tr>
<td>sp</td>
<td>P81760</td>
<td>TL17_ARATH</td>
</tr>
<tr>
<td>tr</td>
<td>Q92160</td>
<td>11.6_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q92VL6</td>
<td>18.3_kDa_protein</td>
</tr>
<tr>
<td>sp</td>
<td>Q9W33</td>
<td>19.9_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9LVV5</td>
<td>15_kDa_protein</td>
</tr>
<tr>
<td>tr</td>
<td>Q9FL23</td>
<td>proteinase_D1</td>
</tr>
<tr>
<td>tr</td>
<td>Q9SSA5</td>
<td>40_kDa_protein</td>
</tr>
</tbody>
</table>
Proteome map of the chloroplast lumen of Arabidopsis thaliana.

Maria Schubert, Ulrika A. Petersson, Brian J. Haas, Christiane Funk, Wolfgang P. Schröder, and Thomas Kieselbach

The preparation of lumen content from Arabidopsis chloroplasts (page 8355) should be completed as follows. Arabidopsis chloroplasts were prepared from 100 g of leaves (wet weight) as described (1). The leaves were divided in portions of 20 g that were blended five times for 1 s using a Heidolph DIAX 900 homogenizer in 170 ml of 20 mM Tricine-NaOH (pH 8.4), 300 mM sorbitol, 10 mM EDTA, 10 mM KCl, 0.25% (w/v) bovine serum albumin, 4.5 mM sodium ascorbate, and 5 mM cysteine. The preparation was continued as described (1), but the chloroplasts were washed and resuspended in 20 mM Hepes-NaOH (pH 7.8), 300 mM sorbitol, 5 mM MgCl₂, 2.5 mM EDTA, and 10 mM KCl.

REFERENCE

Co-repressor release but not ligand binding is a prerequisite for transcription activation by human retinoid acid receptor α ligand-binding domain.

Hung-Ying Kao, Chris C. Han, Anton A. Komar, and Ronald M. Evans

Dr. Komar's name was misspelled. The correct spelling is shown above. The online version has already been corrected in departure from print.