The Plant Cell, Vol. 16, 478–499, February 2004, www.plantcell.org © 2004 American Society of Plant Biologists

In-Depth Analysis of the Thylakoid Membrane Proteome of *Arabidopsis thaliana* Chloroplasts: New Proteins, New Functions, and a Plastid Proteome Database [W]

Giulia Friso,[a,1] Lisa Giacomelli,[a,1] A. Jimmy Ytterberg,[a,1] Jean-Benoit Peltier,[a,1] Andrea Rudella,[a] Qi Sun,[b] and Klaas J. van Wijk[a,2]

[a] Department of Plant Biology, Cornell University, Ithaca, New York 14853
[b] Computational Biology Service Unit, Cornell University, Ithaca, New York 14853

An extensive analysis of the *Arabidopsis thaliana* peripheral and integral thylakoid membrane proteome was performed by sequential extractions with salt, detergent, and organic solvents, followed by multidimensional protein separation steps (reverse-phase HPLC and one- and two-dimensional electrophoresis gels), different enzymatic and nonenzymatic protein cleavage techniques, mass spectrometry, and bioinformatics. Altogether, 154 proteins were identified, of which 76 (49%) were α-helical integral membrane proteins. Twenty-seven new proteins without known function but with predicted chloroplast transit peptides were identified, of which 17 (63%) are integral membrane proteins. These new proteins, likely important in thylakoid biogenesis, include two rubredoxins, a potential metallochaperone, and a new DnaJ-like protein. The data were integrated with our analysis of the lumenal-enriched proteome. We identified 83 out of 100 known proteins of the thylakoid localized photosynthetic apparatus, including several new paralogues and some 20 proteins involved in protein insertion, assembly, folding, or proteolysis. An additional 16 proteins are involved in translation, demonstrating that the thylakoid membrane surface is an important site for protein synthesis. The high coverage of the photosynthetic apparatus and the identification of known hydrophobic proteins with low expression levels, such as cpSecE, Ohp1, and Ohp2, indicate an excellent dynamic resolution of the analysis. The sequential extraction process proved very helpful to validate transmembrane prediction. Our data also were cross-correlated to chloroplast subproteome analyses by other laboratories. All data are deposited in a new curated plastid proteome database (PPDB) with multiple search functions (http://cbsusrv01.tc.cornell.edu/users/ppdb/). This PPDB will serve as an expandable resource for the plant community.

INTRODUCTION

Chloroplasts in green algae and higher plants contain photosynthetic thylakoid membranes with four multisubunit protein complexes (photosystem I [PSI], photosystem II [PSII], ATP synthase, and cytochrome *b6f* complexes), each with multiple cofactors. These four complexes are composed of at least 70 different proteins that perform the photosynthetic reactions (Ort and Yocum, 1996; Rochaix et al., 1998; Wollman et al., 1999).

The functional contributions of many of the individual subunits of these complexes are now understood to varying degrees. This understanding was achieved through site-directed mutagenesis, targeted antisense techniques, and insertional mutagenesis with T-DNAs or transposons, often in combination with forward genetics screens (Ort and Yocum, 1996; Leister, 2003). Nevertheless, there are still known subunits with undetermined functions. Sequencing of the *Arabidopsis thaliana* genome has allowed the discovery of paralogues of known photosynthetic proteins; expression, function, and redundancy of some of these paralogue pairs, such as PsaH1 and PsaH2, now have been tested (Varotto et al., 2002). However, there are still quite a few predicted paralogues that have not been experimentally observed.

The biogenesis, maintenance, and regulated breakdown of these photosynthetic complexes are very complex processes. They require coordinated expression of the nuclear and plastid genome and coordination between protein accumulation and biosynthesis of many cofactors in addition to the formation and maintenance of the lipid bilayer. Throughout the last decade or so, significant progress has been made in identifying and characterizing proteins involved in these processes, such as proteases (Chen et al., 2000; Sakamoto et al., 2002), protein translocators of the Sec and twin Arg translocation (TAT) machinery in *Zea mays* (maize) and *A. thaliana* (reviewed in Mori and Cline, 2001), metal cofactor transporters (Burkhead et al., 2003; Shikanai et al., 2003), proteins involved in RNA maturation (Monde et al., 2000), and specific assembly factors (Meurer et al., 1998; Lennartz et al., 2001).

The thylakoid membrane system also must adjust to changes in abiotic conditions. This requires short-term responses, such as state transitions and a buildup of quenching components, as well as long-term responses, such as a change of PSI/PSII ratios (Aro and Andersson, 2001). These regulators are typically

[1] These authors contributed equally to this work.
[2] To whom correspondence should be addressed. E-mail kv35@cornell.edu; fax 607-255-5407.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Klaas J. van Wijk (kv35@cornell.edu.)
[W] On-line version contains Web-only data.
Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.017814.

expressed at much lower levels than components of the photosynthetic apparatus, making their biochemical identification quite challenging. Several regulatory proteins, such as thylakoid kinases, now have been identified via genetic screens (Snyders and Kohorn, 1999; Depege et al., 2003).

The thylakoid membrane system also must be protected against abiotic stresses, of which oxidative stress is one of the most prominent. Oxidative damage results from incomplete detoxification of reactive oxygen species. To prevent and respond to oxidative stress, an antioxidative defense system is expressed in the chloroplast, consisting of proteins and scavenging molecules (Karpinski et al., 1999). Several of these proteins are associated with the thylakoid membrane and include thylakoid ascorbate peroxidase (Yabuta et al., 2002), three m-type thioredoxins (Issakidis-Bourguet et al., 2001), peroxiredoxins (Konig et al., 2002), and superoxide dismutases. These proteins are fairly abundant, with expression levels at two to three orders of magnitude lower than the photosynthetic apparatus (Peltier et al., 2002). If protection of the thylakoid membrane system is not complete, damage to components such as proteins, lipids, and cofactors will occur, promoting the need for replacement.

In addition to the abundant proteins of the photosynthetic apparatus, several nonphotosynthetic thylakoid membrane proteins have been identified, whereas others can be postulated (van Wijk, 2001a). The expression levels of many of these nonphotosynthetic thylakoid proteins are expected to be much lower than the structural proteins of the photosynthetic complexes, posing a challenge to their comprehensive identification.

With the advent of proteomics and mass spectrometry (MS), systematic identification of proteins has become possible, as demonstrated in several studies in different organisms (Tyers and Mann, 2003), including plants (Millar et al., 2001; van Wijk, 2001b; Gomez et al., 2002; Ferro et al., 2003; Watson et al., 2003). However, identification of low abundance proteins will require multistep fractionation, despite the very high sensitivity of modern MS. Membrane proteomes are especially challenging and very few membrane proteomes have been analyzed successfully. Reverse-phase (RP) HPLC and gel filtration of intact proteins using different solvent systems have been used to purify different integral membrane proteins (Tarr and Crabb, 1983; Sussman, 1988; Lew and London, 1997; Zhang et al., 2001; Whitelegge et al., 2002) and to separate abundant members of the thylakoid grana proteome (Gomez et al., 2002). Other groups (Molloy et al., 1999; Seigneurin-Berny et al., 1999; Ferro et al., 2000, 2002, 2003) have explored fractionation with organic solvents, followed by one-dimensional electrophoresis (DE) SDS-PAGE as a way to reduce membrane proteome complexity and to remove soluble proteins from integral membrane proteins. Such organic solvent fractionation has proven useful because it enriches for hydrophobic proteins.

In this study, we analyzed the peripheral and integral thylakoid membrane proteins by a number of complementary approaches to identify both high and low abundance thylakoid proteins. Altogether, 154 proteins were identified, 76 of which (49%) are integral membrane proteins. Many thylakoid-associated and integral membrane proteins that are not part of the photosynthetic complexes were identified; many of them are novel proteins with diverse functions. In recent years, a number of comprehensive proteomics studies have focused on the chloroplast envelope (Ferro et al., 2002, 2003; Froehlich et al., 2003) and the thylakoid lumen (Peltier et al., 2002; Schubert et al., 2002). The experimental data we present here and in our previous study (Peltier et al., 2002) are cross-correlated to these previous proteomics studies.

Our experimental data can be searched, using different parameters and text strings, in a new manually curated plastid proteome database (PPDB) (http://cbsusrv01.tc.cornell.edu/users/ppdb/). PPDB is a specialized proteome database dedicated to plant plastids. The main objective is to provide a centralized data deposit for both predicted and experimentally identified plastid proteins, their annotated functions, as well as their molecular and biophysical properties. The content of PPDB can be accessed directly through its Web interface. Multiple search options are provided so that the user can retrieve information based on gene identification number, functional annotation, or various protein properties. Active links to other databases (e.g., The Arabidopsis Information Resource [TAIR]) are present.

This study serves three objectives: (1) to provide a comprehensive experimental characterization of the thylakoid proteome, identifying new proteins and paralogues of the photosynthetic apparatus as well as proteins that potentially play a critical role in thylakoid biogenesis and function; (2) to add proteins to the set of experimentally identified plastid proteins that can be used to further improve prediction of the plastid proteome; and (3) to introduce the PPDB to the public domain.

RESULTS

General Overview of Subproteome Purification and Analysis

The peripheral and integral thylakoid membrane proteomes were extracted into complementary protein populations and separated by 1- or 2-DE gels or by RP-HPLC, as depicted in Figure 1A. Such extensive fractionation was undertaken to adjust methods of protein separation to hydrophobicity and to determine the localization and membrane affinity for each protein. This will aid in the identification of proteins with low relative expression levels, provide information relevant to function, and validate transmembrane (TM) prediction. In addition, we used different strategies to improve the success rate of identification by MS, as discussed in greater detail in Methods.

Three complementary subproteomes were generated, including (1) a peripheral-enriched subproteome. Isolated thylakoids from *A. thaliana* leaves were subjected to Yeda press treatment to remove most soluble lumenal proteins and associated stromal proteins, similarly as in Peltier et al. (2002). The peripheral thylakoid proteins then were extracted from these Yeda press–treated thylakoid vesicles by 0.1 M Na_2CO_3 under repetitive sonication. This released remaining lumenal proteins as well as proteins that interact tightly with the thylakoid membrane via electrostatic interactions. Extracted proteins were separated on 2-DE denaturing gels with immobilized pH gradient (IPG) strips in the first dimension, followed by

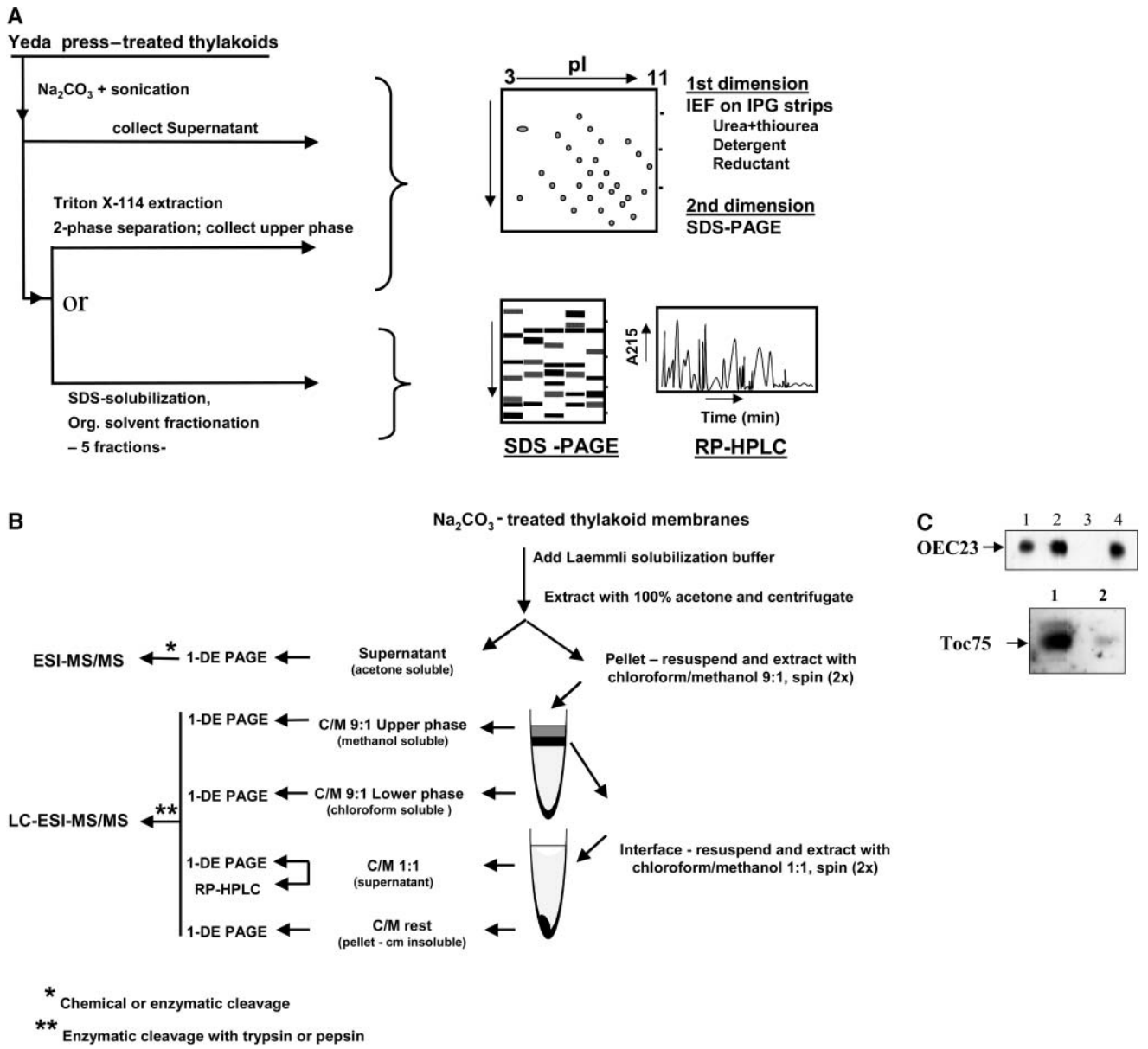

Figure 1. Overview of the Generation of the Three Thylakoid Subproteomes and Their Protein Separation and Identification by MS.

(A) Sequential extraction of thylakoid membrane proteins based on membrane affinity and solubility in organic solvents. Yeda press–treated purified thylakoid membranes were incubated with Na_2CO_3, and released proteins were collected. Residual peripheral thylakoid proteins in these Na_2CO_3-stripped thylakoid membranes were extracted by incubation with Triton X-114, followed by two-phase separation. The water-soluble proteins in the upper phase were collected. These two water-soluble proteomes were separated on 2-DE gels, with IPG strips in the first dimension, followed by SDS-PAGE in the second dimension. To analyze the hydrophobic thylakoid proteins, the Na_2CO_3-stripped thylakoids were solubilized in SDS and sequentially extracted with organic solvents, as explained in **(B)**. The different fractions were separated by RP-HPLC or by SDS-PAGE.

(B) Scheme of fractionation and identification of the hydrophobic membrane proteome by sequential organic solvent extraction, followed by 1-DE PAGE or RP-HPLC and identification by MS analysis. Thylakoid membranes were stripped of lumenal and peripheral proteins with chaotropic agents and sonication and dissolved in Laemmli buffer before organic solvent extraction. Membranes first were dissolved in Laemmli solubilization buffer, followed by extraction in 100% acetone to remove the large amount of chlorophylls and carotenoids. These chromophores would otherwise interfere with further analysis. The remaining protein population, mostly devoid of pigments, was further fractionated by extraction with a mixture of C/M in a ratio of 9:1. This resulted in a three-phase separation with one subset of proteins fractionating to the upper, methanol-enriched phase, one set of proteins dissolving in the lower, chloroform-enriched phase, and a fraction of proteins accumulating in the interphase. The interphase was collected and further extracted by a C/M mixture in a ratio of 1:1, which yielded a subset of C/M 1:1 soluble proteins and a C/M 1:1 insoluble pellet (rest).

(C) Protein gel blot analysis. Partitioning of the lumenal peripheral protein OEC23 was assessed by protein gel blot analysis. The presence of chloroplast envelopes in the starting material for the proteome analysis in this study was determined by blotting with Toc75, a well-known envelope protein. Lane 1, intact chloroplasts; lane 2, Yeda press–treated thylakoid membranes; lane 3, Na_2CO_3-stripped, Yeda press–treated thylakoids; lane 4, Na_2CO_3-extracted thylakoid proteins. Thirty micrograms of chlorophyll was loaded for lanes 1, 2, and 3. Lane 4 contained Na_2CO_3-extracted proteins from Yeda press–treated thylakoids equivalent to 30 μg of chlorophyll.

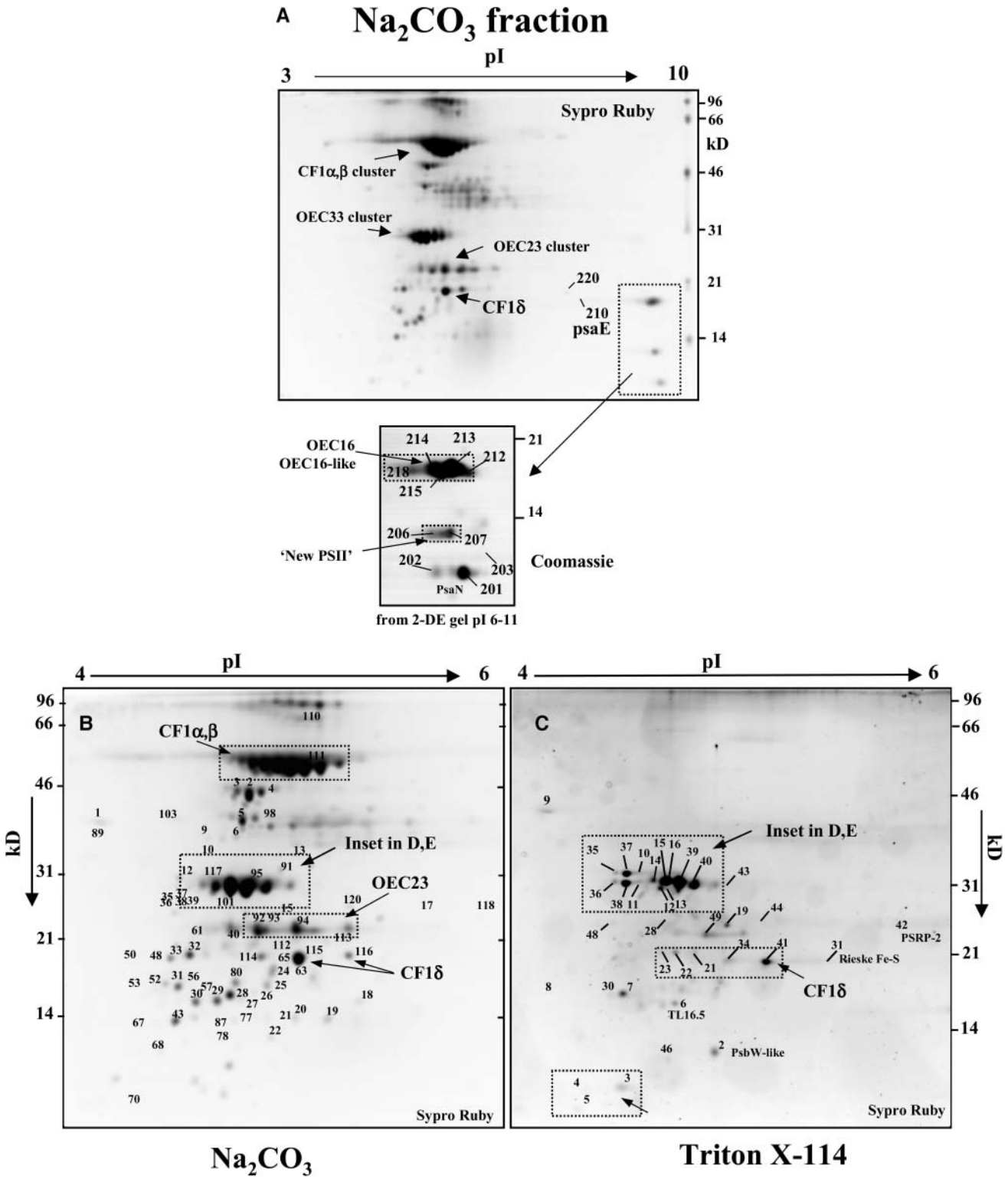

Figure 2. 2-DE Gels of the Na_2CO_3 and Triton X-114–Extracted Thylakoid Proteomes, Stripped by Yeda Press.

Data from the gel identifications are summarized in Table 1 and in Table 1 in the supplemental data online. Details of experimental coordinates and MS identification can be found in Tables 2 and 3 in the supplemental data online.

(A) Analysis of the Na_2CO_3-extracted thylakoid proteome stripped of lumenal proteins by three cycles of Yeda press. Sypro Ruby–stained 2-DE gel (pI range 3 to 10 and inset 6 to 11) of peripheral proteins extracted by Na_2CO_3.

(B) and **(C)** Comparison between the Na_2CO_3 **(B)** and Triton X-114 **(C)**–extracted proteomes and the identified proteins in the pI range of 4 to 6.

Tricine-SDS-PAGE in the second dimension, and were identified by MS (Figures 2A and 2B). (2) A peripheral proteome interacting very tightly with the thylakoid membrane was generated. To further extract proteins that interact very tightly with the membrane but that are typically not integral membrane proteins, the remaining salt-extracted, sonicated thylakoid vesicles were incubated in Triton X-114, followed by two-phase separation. This resulted in an upper detergent-depleted phase and a lower detergent-enriched phase. The upper phase, containing the water-soluble proteins, was collected and the proteins were separated on 2-DE denaturing gels and identified by MS (Figure 2C). The lower phase was discarded. (3) The integral membrane proteome, separated into five fractions, also was generated. Integral membrane proteins were extracted from the Na_2CO_3-treated thylakoids. After dissolving these treated thylakoid membranes in SDS, the integral membrane proteins were sequentially extracted, first with acetone and then by chloroform/methanol (C/M) mixtures of increasing polarity. This resulted in a total of five fractions (Figure 1B). Acetone also removed interfering chromophores (chlorophylls and carotenoids). These five fractions of hydrophobic proteins were separated by RP-HPLC or by 1-DE Tricine-SDS-PAGE, and proteins identified by MS.

Table 1 lists all 154 identified proteins in this study and also indicates in which fraction(s) each protein was identified. As reported in Table 1, we cross-checked these proteins to the data in our previous study on the Yeda press–extracted thylakoid proteome, enriched for the lumenal proteins (Peltier et al., 2002), as well as other studies describing a proteome analysis of the thylakoid lumen (Schubert et al., 2002) and the chloroplast envelope (Ferro et al., 2002, 2003; Froehlich et al., 2003). Details for all of the identified proteins in this and our previous study on the lumen (Peltier et al., 2002) and cross-references (Ferro et al., 2002, 2003; Schubert et al., 2002; Froehlich et al., 2003) are summarized in Table 1 in the supplemental data online. This includes protein names, functional domain prediction (by Pfam-HMM at the conservative cutoff 0.1), AGI locus number, different physical-chemical parameters (length, pI value, and grand average of hydropathy value [GRAVY] index as a measure of hydrophobicity; Kyte and Doolittle, 1982) calculated for processed proteins, curated and predicted number of TMs, and localization prediction by a combination of TargetP and LumenP (Westerlund et al., 2003) with additional filters. These data also can be extracted from the PPDB.

Purity of the Starting Material

The purity of the starting material (Yeda press–treated thylakoids) was carefully considered. From our previous extensive MS analysis (Peltier et al., 2002), we concluded that purified thylakoids (on Percoll gradients) have little contamination from nonchloroplast components (e.g., mitochondria, cytosol, and endoplasmic reticulum). Nevertheless, we further improved the purity of the isolated thylakoid by reducing g values for collection of the thylakoids and by slightly altering the Percoll fractionation step. Indeed, very little mitochondrial contamination (only four abundant inner membrane proteins) and a few other (potential) contaminations were found in the analysis in this study (Table 1). Our main concern in this analysis is contamination with envelope proteins because robust criteria are still lacking to distinguish between inner envelope and thylakoid proteins based on amino acid sequence alone. To reduce potential envelope contamination, the thylakoid membranes were washed with a buffer free of Mg^{2+} to stimulate release of interacting envelopes. Starting material (Yeda press–treated thylakoids) was analyzed by protein gel blots to assess contamination from chloroplast envelopes (Figure 1C). Proteins from intact chloroplasts and Yeda press–treated thylakoid membranes were separated by SDS-PAGE and then transferred to blots and probed with Toc75, a well-known component of the chloroplast envelope membrane (Jackson-Constan and Keegstra, 2001). A clear signal from Toc75 was visible in the intact chloroplasts but not in the thylakoid fraction. Overexposure showed a very weak band for

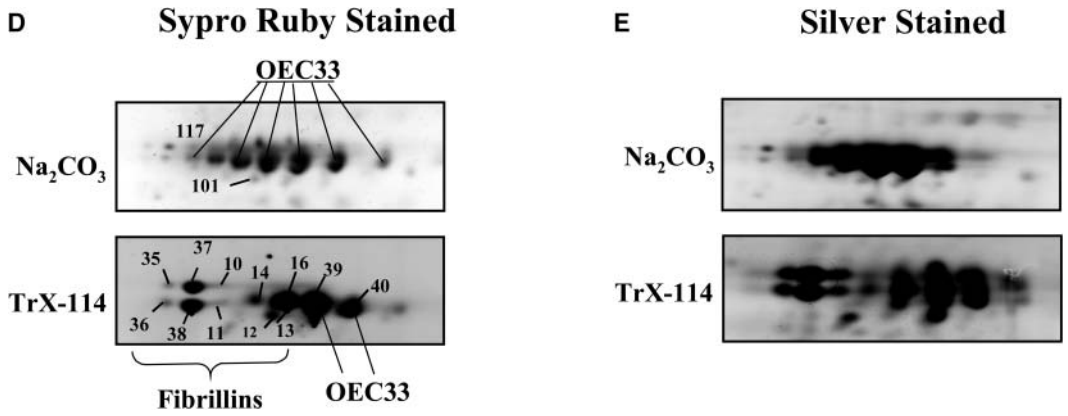

Figure 2. (continued).

(D) and **(E)** Close-ups of the region around the OEC33 cluster of gels stained by Sypro Ruby **(D)** or by silver nitrate **(E)**. The region around this cluster contains a number of fibrillins and they are partially masked by OEC33 and more clearly visible in the Triton X-114 map. Sypro Ruby staining gives superior detection because saturation of staining, as seen in the silver stained maps, is avoided.

Table 1. Overview of All 154 Identified Proteins in This Study

Name and Functional Information	AGI	Triton X-114 2-DE	Na_2CO_3 2-DE	Acetone 1-DE	C/M 9:1 L 1-DE	C/M 9:1 U 1-DE	C/M 1:1 1-DE	C/M rest 1-DE	C/M 9:1 U+L 1-DE	C/M 1:1 RP-HPLC	Ref 1	Ref 2	Refs 3,4	Ref 5
MITOCHONDRIAL (PREDICTED OR EXPERIMENTAL)														
Similar to ATP synthase D chain	At3g52300	21-23							x					
Cytochrome c oxidase subunit 6b	At1g22450		89						x					
Cytochrome c oxidase sub Vb	At3g15640									x				
DegP11 protease	At3g16540									x				
NO PREDICTED cTP OR mTP														
Expressed protein	At4g39730		52,53											
OEP6	At3g63160					x							x,#	x
OM24	At3g52230					x							#	x
Fatty acid elongase-like prot	At3g23840									x				
Organic anion transporter-like p	**At3g20660**									**x**				
Carbonic anhyd homol	**At4g20990**									**x**				
cTP 'STROMA'														
RUBISCO activase	At2g39730							x	x				x,#	x
lipoxygenase AtLOX2	At3g45140					x								x
Cpn21	At5g20720	44									Table 1	x		
30 S ribsomal protein S6	At1g64510		53											x
30S Ribosomal p S5 Isolog L13	At2g33800								x		Table 2			
50S ribosomal protein L12-A	At3g27830	7(1),30(8)	31(3),32(4),52(5), 53(6),56(7)			x(1)			x(1)					
50S ribosomal protein L12-C	At3g27850	7(1),30(8)	31(3),32(4),52(5), 53(6),56(7)			x(1)		x	x(1)		Table 2			
RNA binding protein CP29 B′	At2g37220	48	35-39								Table 2			x
RNA binding protein CP31	At4g24770										Table 2			x
PSRP-2 small	At3g52150	42	17,20,118,120								Table 2			x
PSRP-3 small	At1g68590		18											
PSRP-5 large	At3g56910	8												x
NEW cTP PROTEIN														
Putative protein - also in CN-PAGE -stroma	At5g14910		44						x					
Similar to unknown protein (pir||S75584)	At5g58260						x							x
Putative protein	At4g13200					x		x						
Expressed protein	At5g42765					x								
Expressed protein - Rieske [2Fe-2S] domain	At1g71500							x						x
Expressed protein	At2g26340					x								
ycf37 Synechocystis -stromal predicted	At2g23670	3,4,5	70			x								
Stromal predicted - DnaJ central domain (4 repeats)	At2g34860				x									
Expressed protein	At5g44650		32						x					
Thylakoid phosphoprotein	At3g47070		203		x	x								x
Expressed protein	**At2g46820**				**x**	**x**	**x**	**x**						**x**
Expressed protein - Rhodanese-like domain	**At4g01050**					**x**		**x**						**x**
Putative protein	**At5g55710**						**x**							
Putative protein	**At4g22890**							**x**						
Expressed protein	**At1g74730**						**x**	**x**						
Putative protein - Synech.pir|| S76575 - YGGT family	**At5g36120**				**x**									
Expressed protein	**At4g01150**				**x**	**x**	**x**	**x**						**x**
Expressed protein	**At1g52220**					**x**							**x**	
Putative protein	**At3g51510**					**x**								
Expressed protein	**At3g17930**					**x**								
Hypothetical protein	**At1g73885**					**x**								
Putative protein	**At5g23060**							**x**						
Expressed protein	**At5g08050**					**x**	**x**	**x**						
Hp 17	**At1g42960**					**x**							**#**	
Expressed protein - Rubredoxin and PDZ domain	**At5g17170**							**x**						**x**
Putative protein	**At5g16660**					**x**								**x**
Rubredoxin - homolog in Synechococcus	**At1g54500**					**x**			**x**					

(Continued)

Table 1. (continued).

Name and Functional Information	AGI	Triton X-114 2-DE	Na_2CO_3 2-DE	Acetone 1-DE	C/M 9:1 L 1-DE	C/M 9:1 U 1-DE	C/M 1:1 1-DE	C/M rest 1-DE	C/M 9:1 U+L 1-DE	C/M 1:1 RP-HPLC	Ref 1	Ref 2	Refs 3,4	Ref 5
PHOTOSYSTEM II														
psbA D1	**Atcg00020**					**x**	**x**	**x**						
psbB CP47	**Atcg00680**					**x**	**x**	**x**	**x**					
psbC CP43	**Atcg00280**					**x**	**x**	**x**	**x**					
psbD D2	**Atcg00270**					**x**	**x**	**x**	**x**	**x**				
psbE cytb559a	**Atcg00580**				**x**	**x**	**x**	**x**	**x**	**x**				
psbF cytb559b	**Atcg00570**			**x**	**x**	**x**				**x**				
psbH-phospho	**Atcg00710**				**x**	**x**	**x**		**x**	**x**				
psbL	**Atcg00560**			**x**										
psbO OEC33	At5g66570	10,13-16(25), 39,40	12(1),13(2),91, 95,117			x(2)		x	x(2)		Table 1	x		
psbO OEC33-like	At3g50820	10,13-16(25)	12(1),13(2)			x(2)		x	x(2)		Table 1	x		
psbP OEC23 Tat ITP	At2g30790								x(3)			x		
psbP OEC23 Tat ITP	At1g06680		92-94						x,x(3)		Table 1	x		x
OEC23-like Tat ITP	At4g15510	19									Table 1	x		
OEC23-like Tat ITP	At2g39470							x			Table 1			
OEC23-like Tat ITP	At1g76450		65								Table 1	x		
psbQ OEC16 Tat ltp	At4g21280		212-215,218			x					Table 1	x		
psbQ OEC16-like Tat ITP	At4g05180		212,215			x			x		Table 1	x		
PsbQ domain TAT ITP	At1g14150					x								
PsbQ domain Tat ITP	At3g01440					x								
psbR	**At1g79040**					**x**		**x**						
psbS	**At1g44575**				**x**	**x**	**x**	**x**						
psbT/ycf8	**Atcg00690**									**x**				
psbTn-2	At1g51400					x			x					
psbW - 'LTP' but spontaneous	**At2g30570**			**x**										
psbW -like	At4g28660	2												
psbX- ITP but spontaneous	**At2g06520**			**x**						**x**				
psbY1+2	**At1g67740**									**x**				
psbZ'ycf9	**Atcg00300**			**x**										
PSII Lumen Tat ITP	At1g03600		68,206,207			x		x			Table 1			
LHCII PROTEINS														
LHCII-1.1 - 100% identical to lhcb-1.2,3-mature	**At1g29910**				**x(1)**	**x(3)**	**x(1)**	**x(5)**	**x(4)**	**x(1)**				
LHCII-1.2 - 100% identical to lhcb-l.l,3-mature	**At1g29920**				**x(1)**	**x(3)**	**x(1)**	**x(5)**	**x,x(4)**	**x(1)**				
LHCII-1.3 - 100% identical to lhcb-1.1,2 -mature	**At1g29930**				**x(1)**	**x(3)**	**x(1)**	**x(5)**	**x(4)**	**x(1)**				
LHCII-1.4	**At2g34430**				**x(2)**	**x,x(2)**	**x,x(2)**	**x,x(2)**	**x(4)**	**x(2)**			x	x
LHCII-1.5	**At2g34420**				**x(2)**	**x(2)**	**x(2)**	**x(2)**	**x(4)**	**x(2)**				
LHCII-2.1	**At2g05100**				**x(3)**	**x(5)**	**x(3)**	**x(7)**	**x(4)**					
LHCII-2.2	**At2g05070**				**x(3)**	**x(5)**	**x(3)**	**x(7)**	**x,x(4)**					
LHCII-2.4	**At3g27690**				**x(3)**	**x(5)**	**x(3)**	**x(7)**	**x,x(4)**				x	
LHCII-3	**At5g54270**				**x**	**x**	**x**	**x**	**x**	**x**				
LHCII-4.1-CP29	**At5g01530**				**x**	**x**	**x**	**x**	**x**					
LHCII-4.2 - CP29	**At3g08940**				**x**	**x**	**x**	**x**						x
LHCII-4.3 -CP29	**At2g40100**					**x**	**x**	**x**						
LHCII-5 - CP26	**At4g10340**				**x**	**x**	**x**	**x**	**x**	**x**				
LHCII-6 - CP24	**At1g15820**				**x**	**x**	**x**	**x**	**x**	**x**			x	
cytb6f COMPLEX														
petA - cytochrome f (cleavable LTP of 35 aa)	**Atcg00540**					**x**	**x**	**x**						
petB - Cytochrome b6	**Atcg00720**				**x**		**x**	**x**	**x**					
petC - Rieske Fe-S protein - LTP	At4g03280	1#,31,32#				x	x	x					x	
petD - subIV	**Atcg00730**				**x**		**x**		**x**	**x**				
ELECTRON TRANSPORT														
FNR-1	At5g66190		10,89			x		x			Table 2	x	#	x
FNR-2	At1g20020							x		x				
PHOTOSYSTEM I														
psaA - subunit Ia	**Atcg00350**					**x**								
psaB - subunit Ib	**Atcg00340**					**x**	**x**	**x**						
psaC - subunit VII - stromal side	Atcg01060				x	x		x	x					
psaD-1 subunit II - stromal side-tight	At1g03130							x(1)			Table 2			x
psaD-2 subunit II - stromal side-tight	At4g02770							x(1)		x				

(Continued)

Table 1. (continued).

Name and Functional Information	AGI	Triton X-114 2-DE	Na_2CO_3 2-DE	Acetone 1-DE	C/M 9:1 L 1-DE	C/M 9:1 U 1-DE	C/M 1:1 1-DE	C/M rest 1-DE	C/M 9:1 U+L 1-DE	C/M 1:1 RP-HPLC	Ref 1	Ref 2	Refs 3,4	Ref 5
psaE-1 subunit IV - stromal side	At4g28750		210		x(4)	x	x	x			Table 2			
psaE-2 subunit IV - stromal side	At2g20260				x(4)	x		x						
psaF- subunit III - LTP - hydrophobic	**At1g31330**				**x**	**x**	**x**	**x**	**x**					**x**
psaG - subunit V - stromal side	**At1g55670**				**x**	**x**	**x**	**x**	**x**	**x**			**x**	**x**
psaH-1 - subunit VI	**At3g16140**				**x,x(5)**	**x**	**x**	**x(8)**	**x**	**x**				
psaH-2 - subunit VI	**At1g52230**				**x(5)**	**x**	**x**	**x(8)**		**x**				
psaJ - subunit VIII	**Atcg00630**			**x**						**x**				
psaK - subunit X	**At1g30380**						**x**		**x**	**x**				
psaL - subunit XI (also named V)	**At4g12800**				**x**		**x**	**x**	**x**	**x**			**x**	
psaN -TAT LTP	At5g64040		201,202		x	x	x	x	x		Table 1			
LHCI-1-1 - LHCI-730	**At3g54890**				**x**	**x**	**x**	**x**	**x**				**x**	
LHCI-2.1 - LHCI-680B	**At3g61470**		**x**			**x**	**x**			**x**				
LHCI-3 - LHCI-680A CAB4	**At1g61520**				**x**	**x**	**x**	**x**	**x**	**x**			**x**	
LHCI-4 - LHCI-730	**At3g47470**				**x**	**x**	**x**	**x**	**x(5)**	**x**			**x**	
LHCI-5 - new	**At1g45474**								**x(5)**					
CFO-CF1														
CFO-I - atpF	**Atcg00130**				**x**	**x**	**x**	**x**						
CFO-II - atpG	**At4g32260**	**30**	**56**		**x**	**x**	**x**	**x**	**x**					**x**
CFO-III - atpH	**Atcg00140**			**x**	**x**					**x**				
CFO-IV - atpI	**Atcg00150**						**x**							
CF1b - atpB	Atcg00480		122-126					x,x(3)	x		Table 2	x		
	Atcg00120		110,111,122-125		x	x		x			Table 2	x		
CF1a - atpA														
CF1e - atpE	Atcg00470		20,87		x	x	x	x			Table 2			
CF1y - atpC	At4g04640		89			x		x			Table 2			x
CF1d - atpD	At4g09650	31,41	32,33,114-116		x	x	x	x	x		Table 2			x
ALTERNATIVE ELECTRON FLOW														
NDH E (4L)	**Atcg01070**				**x**		**x**							
NDHI	Atcg01090							x						
PGR5	At2g05620				x									
TRANSLOCATION AND ASSEMBLY														
SecE	**At4g14870**													
HCF136 Tat Itp	At5g23120		6,98,103,110			x					Table 1	x		
STRESS DEFENSE AND OTHERS														
Fibrillin	At2g35490	9	1					x			Table 2			
Fibrillin	At3g23400	12-15,39						x			Table 2	x		
Fibrillin	At4g22240	11,36,38						x(9)			Table 2	x		
Fibrillin	At4g04020	10,35,37	12					x,x(9)			Table 2	x		
Fibrillin	At3g26070							x(2)						
Fibrillin	At3g26080							x(2)						
Thioredoxin m2	At4g03520	46	68,78								Table 1			
Thioredoxin m4	At3g15360		77,87								Table 1			x
Asc-perox lumen Tat Itp	At4g09010							x			Table 1	x		
Ferritin 1	At5g01600	19												
Ohp1 or Lil2/Hlip/Scp	**At5g02120**				**x**									
Ohp2	**At1g34000**				**x**	**x**	**x**							
THYLAKOID PROTEASES AND PEPTIDASE														
FtsH2 (Var2)	**At2g30950**		**50*,68***					**x(4)**				**x**		**x**
FtsH8	**At1g06430**							**x(4)**						**x**
FtsH5	**At5g42270**							**x**						**x**
NEW LUMENAL PROTEINS AND ISOMERASES														
Isomerases TAT ITP	At3g10060		53								Table 1			
Isomerases	At4g39710							x			Table 1			
Isomerases	At3g15520		6,9									x		
Thylakoid lumen 18.3 kD -ambiguous - Duf477	At1g54780					x		x			Table 1	x		
Thylakoid lumen protein TL16.5	At4g02530	6	26,28,57,80			x					Table 1	x		
Pentapeptide repeat TL17.4	At5g53490		25								Table 1	x		
Thylakoid lumen pentapeptide repeat	At1g12250		112,113											

Proteins are classified per location or function. Indicated is in which fraction(s) the proteins were identified. Details concerning identification, functional domains, and other parameters can be found in the Tables 1, 2, and 3 in the supplemental data online and in the PPDB. Proteins (76 in total) indicated in bold text are integral membrane proteins. Numbers in parentheses indicate that some of the MS data match to a paralogue; these paralogue pairs are connected by the same integer within each column. The explanation for each fraction is self-explanatory. In addition, the identified proteins were cross-checked with our recent proteome study on the thylakoid-lumenal enriched proteome from Peltier et al. (2002) (ref. 1), the study on the lumenal proteome from Schubert et al. (2002) (ref. 2), and studies of the *A. thaliana* choroplast envelope from Ferro et al. (2002, 2003) (refs. 3 and 4) and from Froehlich et al. (2003) (ref. 5).

Toc75 in the thylakoids, which was estimated to be <0.5% of the band in the intact chloroplasts (Figure 1C).

In the extensive MS analysis in this article, we did not find any of the well-known components of the protein import apparatus (the Tic and Toc complex; reviewed in Jarvis and Soll, 2002) nor the very abundant triose phosphate translocator of the inner membrane (Schneider et al., 2002) or other known envelope transporters (Flugge, 1998; Fischer and Weber, 2002). Therefore, we concluded that the starting material for the thylakoid proteome analysis contained very little contamination from the chloroplast envelopes. This was not surprising because thylakoid membranes are much more abundant than envelopes and because thylakoids sediment at much lower g values than envelopes. Nevertheless, all of our data were cross-correlated with several published envelope proteome articles (Table 1; Table 1 in the supplemental data online), and we comment on the overlapping identities.

Finally, we verified the presence of peripheral proteins in the Na_2CO_3-treated membranes by protein gel blots with OEC23, a PSII component associated with the lumenal side of the thylakoid membrane. Intact chloroplasts, Yeda press–treated thylakoids, Na_2CO_3-stripped thylakoids, and the extracted peripheral proteins were probed. This showed that OEC23 levels were below the detection limit in the Na_2CO_3-treated membranes. This was confirmed by the 2-DE gel of the Triton X-114 population because the OEC23 levels were extremely low (Figure 2C).

Separation and Identification of the Peripheral Thylakoid Proteome

Figure 2A shows a wide-range pI gradient 2-DE gel (pI 3 to 10) of the Na_2CO_3-extracted proteome as well as a region of a 6 to 11 pI gel with spot numbers. Figures 2B and 2C show 2-DE gels of the Na_2CO_3 and Triton X-114 proteomes with spot numbers (separate set of numbers for each fraction) corresponding to entries in Table 1. Experimental pI, molecular weight, and MS data can be found in Tables 2 (Na_2CO_3) and 3 (Triton X-114) in the supplemental data online and also can be extracted from the PPDB. For more details, see Table 1 in the supplemental data online.

The carbonate-extracted proteome map was very much dominated by OEC33, CF1α, and CF1β; together, they represented >60% of the total protein mass of this subproteome (Figures 2A and 2B). Other abundant proteins were OEC23 and OEC16 as well as the smaller members of the CF1 complex (δ,γ, and ε). The Triton X-114–extracted subproteome represented only ~7% of the total peripheral proteome and was dominated by OEC33 and the family of fibrillins (Figures 2C to 2E).

In total, 48 different proteins were identified by a combination of matrix-assisted laser-desorption ionization time of flight (MALDI-TOF) MS and electrospray ionization tandem MS (ESI-MS/MS) (Table 1). Twenty-six proteins were unique for the Na_2CO_3 maps and nine were unique for the Triton maps, whereas 13 were found in both populations. Comparison of the two maps clearly gives indications for membrane affinity; CF1α, CF1β, and OEC23 were virtually completely removed by the Na_2CO_3 treatment, whereas OEC33, the fibrillins, and CF1δ could not be completely extracted, indicating very tight membrane interactions. Of these 48 proteins, 30 were identified earlier in the lumenal-enriched maps (Peltier et al., 2002) (Figure 3A), and 19 also were found in another study concerning the lumenal proteome (Schubert et al., 2002).

New proteins without known function that are found in the Na_2CO_3 population but not in the lumenal or Triton fraction (Peltier et al., 2002) are At3g47070, At5g44650, and At1g12250, containing a pentapeptide repeat domain; At5g14910, with a heavy metal binding domain; and At5g44650, with one predicted TM domain but a low GRAVY value (−0.6). Interestingly, At5g14910 also was found in an ~140-kD stromal complex (J.-B. Peltier and K.J. van Wijk, unpublished data) and will be further evaluated in the Discussion. At3g47070 was just recently reported to be a peripheral thylakoid protein (without known function) that can undergo light-induced phosphorylation (Carlberg et al., 2003) and also was found in the analysis of envelopes, possibly as a contaminant (Froehlich et al., 2003).

Proteins that were unique for the Triton X-114 map and were not on the other 2-DE maps included a novel psbW-like protein (At4g28660), the plastid-specific ribosomal protein (PSRP) PSRP-5 (At3g56910), earlier copurified with chloroplast ribosomes (Yamaguchi and Subramanian, 2000), and the Rieske Fe-S protein of the cytochrome *b6f* complex. The latter cytochrome *b6f* protein also was found as a contaminant in envelopes (Ferro et al., 2002).

In particular, the region around the OEC33 cluster showed that the comigrating family of fibrillins could be much better resolved on the Triton X-114 map than it could on the Na_2CO_3 map (Figures 2D and 2E).

Identification of the Hydrophobic Thylakoid Proteome

The five organic solvent fractions with thylakoid proteins each were separated on Tricine-PAGE gels with 6 M urea and visualized by staining with silver nitrate or Coomassie blue (Figure 4A).

Surprisingly, a set of low molecular weight proteins was extracted with acetone (Figure 4A). The 1-DE gel lane between 3 and 10 kD of the acetone fraction was sliced into five bands and digested with trypsin and chymotrypsin or cleaved by limited acid hydrolysis (Shevchenko et al., 2000) or CnBr (van Montfort et al., 2002) and analyzed by manual nano-ESI-MS/MS. Seven proteins belonging to PSII, PSI, or ATP synthase were identified, including the recently characterized psbZ/ycf9 gene product (Swiatek et al., 2001) as well as psbL, psbW, psbX, and psaJ (Figure 4A). Their common denominator is that they are small (<10 kD) and hydrophobic with high GRAVY values (median 0.81) and have one or two TMs (Table 1 in the supplemental data online).

Altogether, 237 gel pieces were systematically excised from the four 1-DE gel lanes of the C/M fractions, in-gel digested by trypsin, and the peptides extracted. All of the samples then were analyzed by automated MALDI-TOF MS and on-line LC-ESI-MS/MS. We also analyzed the pooled upper and lower phase of the C/M 9:1 fraction loaded together on a 1-DE Tricine-PAGE gel, selecting only visible protein gel bands for analysis. This extensive MS analysis identified 124 different proteins, 67 of which had one or more known or predicted α-helical TMs (Figure 4A, Table 1). Thus, 54% of the identified proteins were integral membrane proteins, many of them without known function.

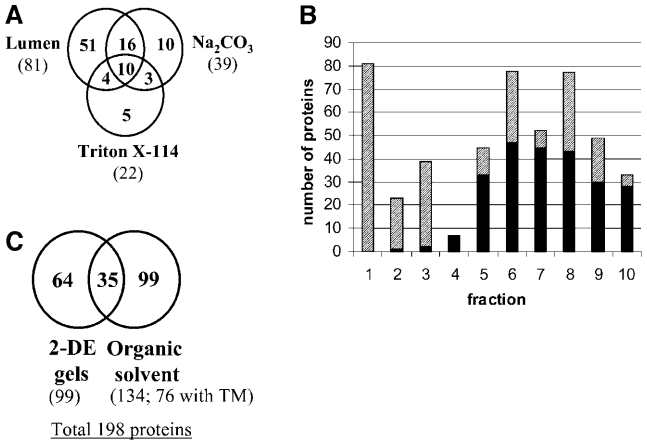

Figure 3. Distribution of Identified Proteins Over the Different Fractions.

(A) Distribution of the identified proteins over the different fractions identified from 2-DE gels. Overlap of the different protein sets is shown.
(B) Stacked bar diagram with the distribution of identified proteins (separated in soluble and TM proteins) over the Yeda press fraction (lumenal enriched), the Na_2CO_3 fraction, the Triton X-114 fraction, and the eight organic solvent fractions. 1, Lumenal-enriched (Yeda press); 2, Triton X-114; 3, Na_2CO_3 fraction; 4, acetone; 5, C/M 9:1 lower phase; 6, C/M 9:1 upper phase; 7, C/M 1:1; 8, residual insoluble pellet; 9, C/M 9:1 total; 10, C/M 1:1 RP-HPLC. Membrane proteins are shown in black.
(C) Distribution of the identified proteins over the 2-DE gels and organic solvent fractions. Overlap between the proteins identified from the 2-DE gels and organic solvents is shown.

These new proteins include two rubredoxins, a potential metallochaperone, and a new DnaJ domain protein. One of these rubredoxins (At1g54500) is a homolog of a rubredoxin in *Synechococcus* sp PCC 7002, which was shown to be involved in the assembly of the interpolypeptide (4Fe-4S) cluster Fx of PSI (Shen et al., 2002a, 2002b). We further review their identities and potential biological functions as well as their membrane affinity in the Discussion.

Finally, RP-HPLC of the C/M 1:1 fraction, in combination with pepsin digestion, was explored as an alternative to gel-based separation and trypsin digestion (Figure 4B). The RP-HPLC fractionation pattern was visualized by Tricine-PAGE gels and showed that there was a distinct, sharp, fractionation pattern (data not shown). The chromatograms were reproducible, and a representative chromatogram is shown in Figure 4B. Each fraction was dried down, digested with pepsin, and analyzed by ESI-MS/MS. Twenty-seven proteins were identified with one or more α-helical TMs and five proteins without predicted (by TMHMM) or known TMs. Some of the proteins have no predicted cTP, and it is not clear if they are truly plastid proteins. The identities of these proteins are briefly evaluated in the Discussion.

In summary, the acetone and C/M extraction of stripped thylakoid membranes and subsequent analysis resulted in the identification of 134 different proteins, of which 76 (57%) had one or more TM. This set was very complementary to the 2-DE analysis, resulting in a total of 198 proteins (Figure 3C). Their functions and biological relevance also are evaluated in the Discussion.

Identification of Closely Related Paralogues

An early analysis of the sequenced *A. thaliana* genome (Arabidopsis Genome Initiative, 2000) showed that only 35% of the predicted proteins are unique in the genome (sequence similarity exceeding a Basic Local Alignment Search Tool [BLASTP] value E $< 10^{-20}$ and extending over at least 80% of the protein length were used as parameters to identify protein families). When proteins within such families are more than ~80% identical, careful verification of the MS results becomes important to determine if only one or more than one paralogues are identified. In some cases, peptides are identified that match to more than one paralogue. Uniquely matched peptides (diagnostic peptides) then are needed to determine which protein is expressed. In each Table, we have indicated if more than one paralogue matches the MS results, using an integer to link the paralogues. Examples of such closely related paralogues (80% or higher sequence identity) that were identified in this study include ribosomal proteins L12A and L12C (two paralogues), OEC33 (two paralogues), OEC23 (two closely related paralogues and many distantly related), psaD (two paralogues), psaE (two paralogues), psaH (two paralogues), and the 15 members of the light-harvesting complex of PSII (LHCII) family. In nearly all cases, we were able to determine which paralogue was identified using a unique or diagnostic peptide. This high success rate of paralogue identification was attributable to the very large number of peptides that were sequenced by MS/MS (well over 1000 or so) and the nature of the fractionation. In the case of the LHCII proteins, we uniquely identified 10 out of 14 proteins. Because the processed LHCII-1.1, LHCII-1.2, and LHCII-1.3 are 100% identical, we could not distinguish them. We found many sequence tags for LHCII-2.1, but they also matched to LHCII-2.2 and LHCII-2.4 (Figure 5). Unique sequence tags were found for LHCII-2.2 and LHCII-2.4 (Figure 5).

Relation between Physical-Chemical Protein Properties and Fractionation/Identification

We analyzed several of the physical-chemical properties of the proteins (GRAVY index, length, and pI value) identified in the different organic solvent fractions, with the goal to understand their fractionation behavior and possibly extrapolate experimental parameters to better predict cellular localization and TM prediction. Physical-chemical properties were calculated for processed proteins, based on cleavage site prediction by TargetP (for the cTPs), LumenP (for the lTPs), and additional filtering (O. Emanuelsson, Q. Sun, and K.J. van Wijk, unpublished data).

Figure 6A shows a cross-correlation between calculated GRAVY values and number or TM per amino acid for all identified membrane proteins. This illustrates that both parameters positively correlated and shows that the identified membrane proteins range (after removal of the cTP/lTP) from a GRAVY value −0.65 to ~1.5 and from 1 TM per 500 amino acids to 1 TM per ~29 amino acids. Figure 6B compares the GRAVY index for the protein population identified in each of the organic solvent fractions identified by 1-DE gels, using a frequency distribution plot. The acetone-extracted proteome clearly separated according

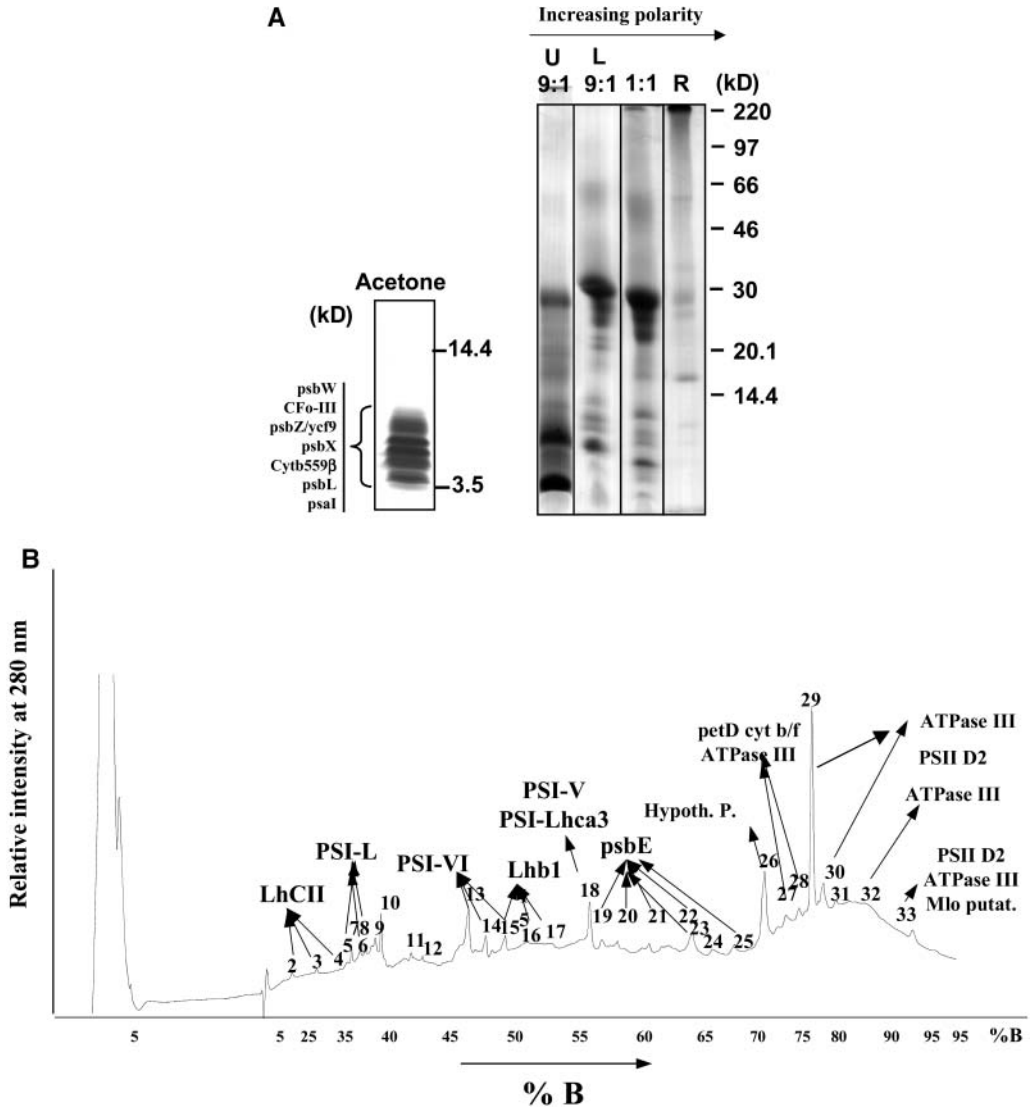

Figure 4. Separation and Identification of the Acetone and C/M-Soluble Thylakoid Proteins.

(A) High-resolution gradient 1-DE Tricine-PAGE of the five protein fractions from sequential organic solvent fractionation of Na_2CO_3-stripped thylakoid membranes. The five fractions are acetone-soluble proteins (panel at left), C/M 9:1 lower phase (9:1 L), C/M 9:1 upper phase (9:1 U), C/M 1:1 soluble proteins (1:1), and the residual insoluble pellet (R). Gels were stained by silver nitrate in the acetone fraction and by Coomassie blue for the other fractions. Proteins identified in the acetone fraction are cytb559β (Atcg00570), psbL (Atcg00560), psbW (At2g30570), psbX (At2g06520), psbZ/ycf9 (Atcg00300), psaI/subunit VIII (Atcg00530), and CFo-III (Atcg00140).

(B) Proteins extracted by C/M 1:1 were separated using RP-HPLC, and a representative chromatogram is shown. Fraction number and the percentage of solvent B are indicated. Protein identities of a number of proteins are indicated. The absorbance was monitored at 280 nm (rather than 214 nm) because isopropanol absorbs strongly at 205 nm; thus, the peaks are not quantitative.

to GRAVY index. The proteins in the other organic solvent fractions show very similar GRAVY distributions.

The Plastid Proteome Database

Proteins identified in this study and our previous study concerning the thylakoid lumen proteome (Peltier et al., 2002) were deposited into a newly developed database, PPDB. Published data concerning the plastid proteome available from other laboratories will be incorporated and used to curate prediction and experimental data. The experimental data are linked to the latest annotation of the *A. thaliana* genome, provided by the Institute for Genomic Research (TIGR) (http://www.tigr.org/tdb/e2k1/ath1/), as well as the mitochondrial and plastid genomes. In addition to the deposited experimental information, the database

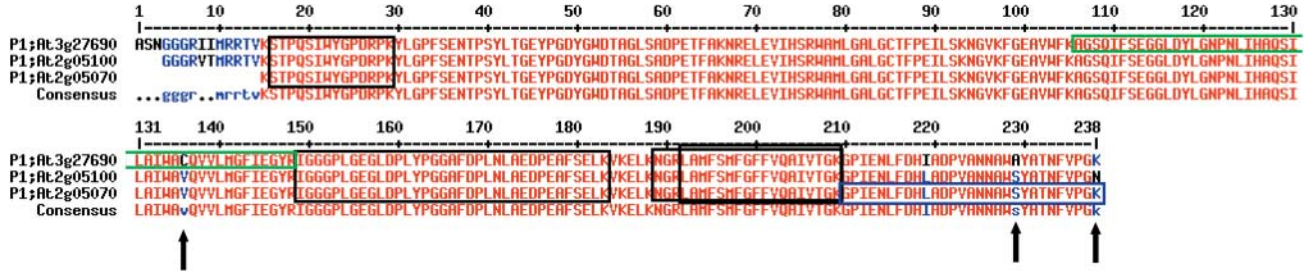

Figure 5. Identification of Paralogues of the LHCII Family.

Alignment and identification of processed LHCII-2.1 (At2g05100), LHCII-2.2 (At2g05070), and LHCII-2.4 (At3g27690) (predicted cTPs were removed). One unique peptide was identified for LHCII-2.2 (boxed in green) and one for LHCII-2.4 (boxed in blue). Many other peptides were identified that matched to all three LHC proteins, some are indicated (boxed in black). Arrows indicate amino acid residues that help to identify the different paralogues.

also provides the predicted localization based on a combination of different predictors and filters (TargetP) (Emanuelsson et al., 2000), combined with a new lumenal proteome predictor (LumenP) (Westerlund et al., 2003) and additional filters for selection of lumenal proteins derived from an experimental data set (Peltier et al., 2002), as well as by prediction of α-helical TM proteins, using TM-HMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Together, this provides predictions of four chloroplast sublocalizations: stroma, membrane (either inner envelope or thylakoid), lumen, and thylakoid membrane with ITP. In the current release (28,581 genes), we predicted 4255 predicted nuclear-encoded chloroplast proteins, divided into 3387 stromal proteins, 291 lumenal proteins (with 71 TAT motifs), 520 TM proteins, as well as 57 proteins that have a predicted ITP and α-helical TM domains, so-called ambiguous thylakoid proteins. Predicted cleavage sites (cTP and lTP) and biophysical properties, including molecular weight, GRAVY, and pI, are provided for each of the predicted proteins, both for their precursor form as well as for their processed form(s) in the case of plastid proteins. It should be noted that despite filtering for a conserved cleavage in proteins without a TAT motif, the lumenal proteome seems somewhat overpredicted (O. Emanuelsson, Q. Sun, and K.J. van Wijk, unpublished data).

A novel Web interface is designed for accessing the PPDB through the World Wide Web. Through this interface, the user can search for information by AGI accession number, functional annotation, and physical properties of all the genes in the PPDB. Details about protein fractionation can be extracted for those proteins that are identified experimentally by our lab. When searching for protein length, molecular weight, and pI, the user can define a range and specify either mature or unprocessed proteins. The display of the search results can be customized by the user. Links to TAIR are provided for each gene.

DISCUSSION

The big challenge in any proteomics project aimed at identifying the members of a proteome is to find proteins of high and low abundance and also to distinguish between individual members of protein families (paralogues). Because molar ratios within proteomes vary over several orders of magnitude, in many cases only the most abundant proteins are identified. This will bias the apparent composition of these proteomes to housekeeping proteins or major structural components. Protein separation typically will improve dynamic resolution. The thylakoid proteome seems particularly challenging because it is so dominated by the photosynthetic apparatus and its hydrophobic chromophores. Therefore, in this study, we used an extensive combination of different fractionation methods and adapted strategies developed by others to identify many of the proteins associated with the thylakoid membrane of chloroplasts of *A. thaliana*. The objectives of this study were to penetrate the thylakoid proteome experimentally as much as possible, such as to identify the photosynthetic apparatus in its (near) completeness and identify known and potential components of the biogenesis apparatus as well as abiotic stress response proteins.

Combining all identified proteins in this study with our previous analysis of the thylakoid lumen (Peltier et al., 2002), 198 proteins were identified, of which 76 (39%) are integral membrane proteins. It is useful to include our complementary study of this lumenal-enriched proteome because it helps to evaluate membrane affinity for each protein and maintain an overview of the thylakoid proteome. Between 21 and 39 of these proteins also were identified in other chloroplast proteomics studies mentioned earlier (Ferro et al., 2002, 2003; Schubert et al., 2002; Froehlich et al., 2003), which can be found in Table 1 in the supplemental data online. Most of these were lumenal proteins, known photosynthetic thylakoid proteins, as well as several abundant stromal proteins, as shown in Table 1 and in Table 1 in the supplemental data online. None of the known proteins of the chloroplast envelope import apparatus or metabolite translocators were identified in our thylakoid membranes, indicative of very low cross-contamination of envelope proteins in our preparations. We want to point out that the envelope preparation analyzed by Froehlich et al. (2003) apparently contained a significant contamination from the thylakoid membrane, as is evident in Table 1 and in Table 1 in the supplemental data online.

To evaluate the success rate of identification of the photosynthetic apparatus, we compared the 198 proteins against all photosynthetic thylakoid proteins. To that end, AGI numbers for all known photosynthetic proteins were collected, and BLAST searches to find additional paralogues were performed. This

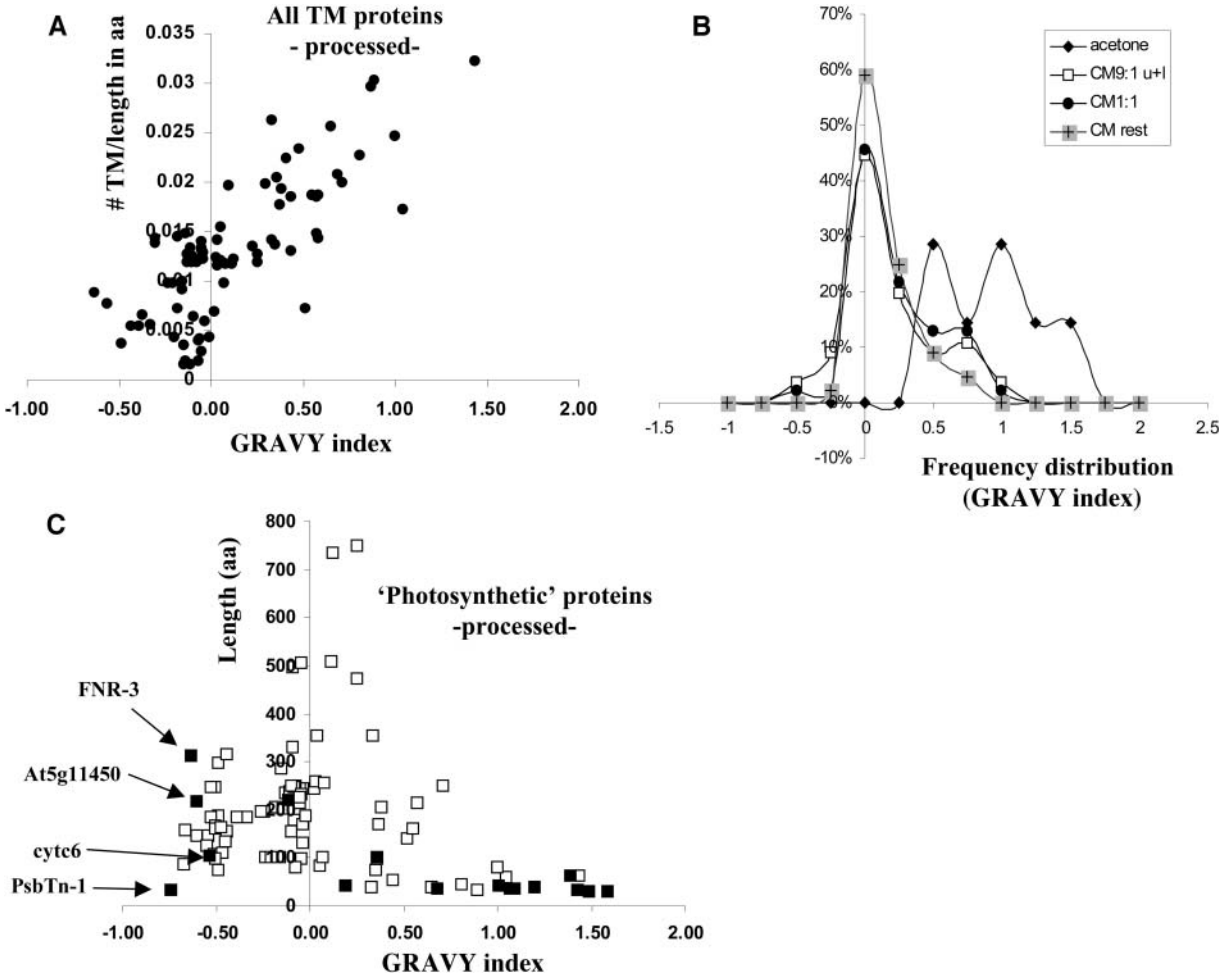

Figure 6. Linking the Physical-Chemical Properties of the Identified Processed (Mature) Proteins to the Sequential Fractionation.

Cleavable presequences were removed.

(A) Scatter plot of GRAVY index and number of TMs per amino acid for all identified proteins with predicted or identified TMs.

(B) Frequency distribution of the GRAVY index for proteins identified in four organic solvent fractions: acetone, C/M 1:9 U+L, C/M 1:1, and the rest fraction.

(C) Correlation between protein length (in amino acids) and GRAVY index for those photosynthetic proteins that were identified (open squares) and those that were not identified (closed squares).

resulted in a list of 99 genes potentially encoding for the photosynthetic apparatus of the thylakoid. We identified 83 of those 99 proteins, as detailed below. Some of the undetected proteins are likely only expressed in nongreen plastids (FNR-3) (Onda et al., 2000) or possibly only under particular conditions (cytochrome *c6*), as in *Chlamydomonas reinhardtii* (Quinn and Merchant, 1998). As we discuss below, we were able to identify an additional ~100 nonphotosynthetic proteins implicated in biogenesis and oxidative stress responses or without known function. Thus, the sequential extraction approach worked well and fulfilled most of our expectations.

Functional Classification of the Thylakoid Proteome

As outlined in Figure 7A and in Table 1 in the supplemental data online, all 198 identified proteins were classified into nine groups. We briefly discuss selected members of each category and comment on their biological function and location.

Thylakoid Photosynthetic Apparatus

Eighty-three proteins within this category were clustered according to the thylakoid protein complex (79) or as electron transport carriers (4). Within each protein complex, proteins were subdivided into integral membrane or soluble proteins (Figure 7B).

For PSII, we collected AGI numbers for 53 putative proteins, including 14 members of the LHCII family; 46 of these 54 proteins (84%) were experimentally identified (Figure 7B). Of the nine PSII proteins that were not identified, six were very small hydrophobic proteins and one was a lumenal OEC23-like protein. Although the PSII complex has been studied intensely, with a high-resolution

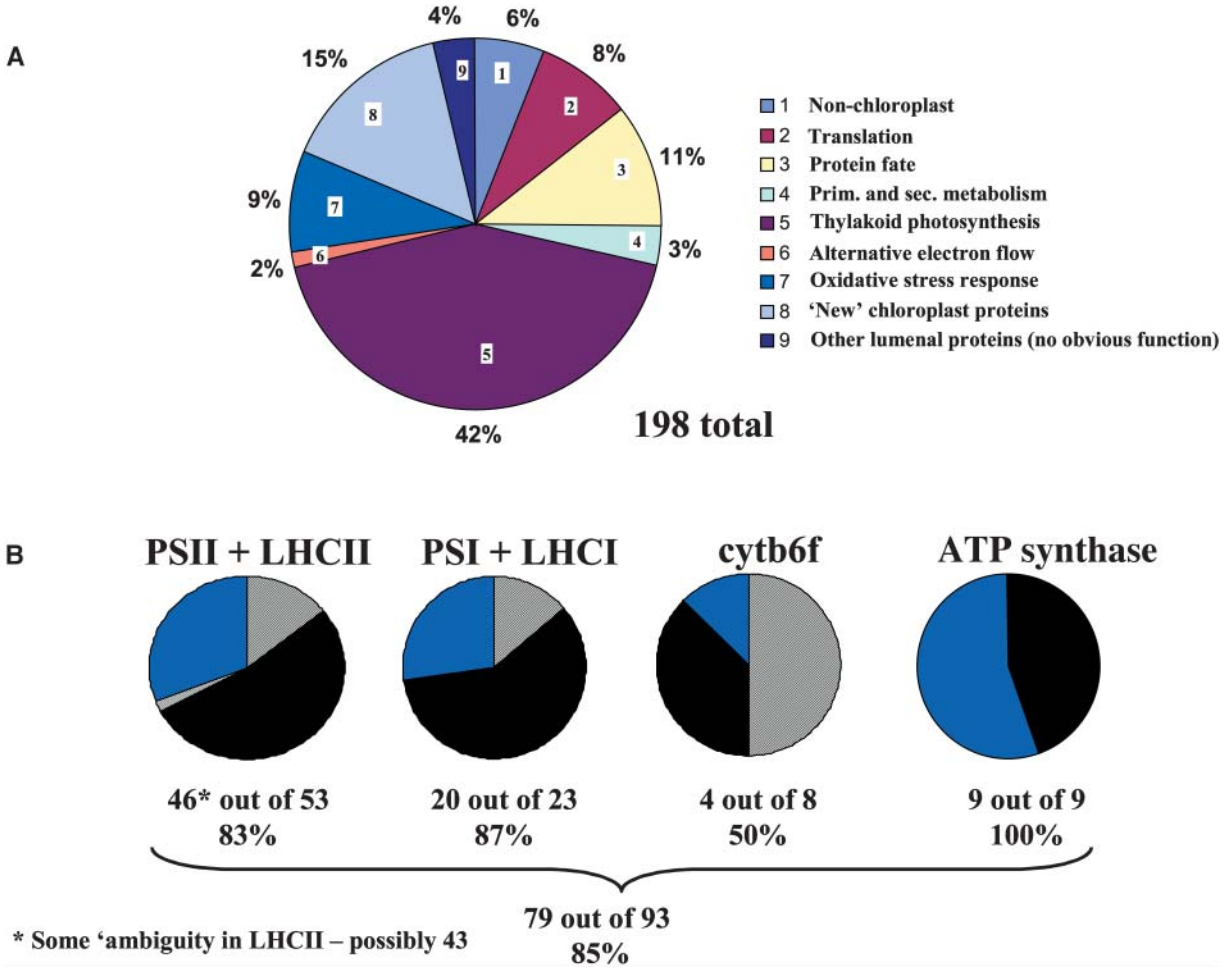

Figure 7. Functional Classification of the Identified Proteins.

(A) Functional classification of all 198 identified proteins. 1, nonchloroplast; 2, translation (ribosomal subunits, RNA binding proteins, and PSRPs); 3, chaperones, protein insertion and assembly, proteases, and protein isomerases; 4, primary and secondary metabolism; 5, thylakoid photosynthetic apparatus (including FNR and PC); 6, alternative electron flow; 7, oxidative stress (Fe-SOD, peroxyredoxins, thioredoxins, ascorbate peroxidase, fibrillins, and Ohp1 and Ohp2); 8, new chloroplast proteins without obvious function; 9, other lumenal proteins without obvious function.
(B) Identified proteins of the photosynthetic machinery, as compared with all predicted proteins and grouped with (black) or without (blue) TM, for PSII (including LHCII), PSI (including LHCI), cytochrome *b6f* complex, and ATP synthase. The striped areas represent the unidentified proteins.

crystal structure obtained for cyanobacterial complexes (Kamiya and Shen, 2003) and cryoelectron microscopy data for higher plant complexes (Nield et al., 2000), there is still a surprising number of distant or closely related paralogues of PSII proteins that are not studied. Here, we highlight some of these putative PSII proteins. We identified two psbW paralogues (At2g30570 and At4g28660). One is well known and hydrophobic with one TM and was identified in the acetone fraction (At2g30570). The other psbW paralogue (psbW-like; At4g28660) was not experimentally identified earlier. It has no predicted TMs and therefore was identified on the 2-DE gel of the Triton X-114 fraction. It is not known if this psbW-like protein is a bona fide PSII protein. We also found a new nuclear-encoded psbT paralogue (psbTn2) of psbTn1 in two different organic solvent fractions. PsbTn1 is a lumenal PSII protein targeted via the TAT pathway. The paralogues are named as such because the chloroplast genome itself encodes for an unrelated protein that also was assigned psbTc or YCF8 (Hankamer et al., 2001). As mentioned in the Results, the identification of some of the LHCII proteins was ambiguous because of the high level of sequence conservation, but most could be identified based on diagnostic peptides. This is of interest because the LHCII composition is regulated by light quality and quantity (Teramoto et al., 2002). It has been very difficult to experimentally distinguish between the subclasses of LHCIIs because this was based on detection with antisera.

We collected AGI numbers for 23 genes encoding for PSI subunits and its antennae; 20 corresponding proteins (86%) were identified (Figure 7B). Those not identified were the membrane proteins psaI (subunit IX), the recently identified psaO gene (Knoetzel et al., 2002), and a light-harvesting complex of PSI protein that could not be well classified (At1g19150). PsaI

is very small (4 kD) and hydrophobic with one TM (Scheller et al., 2001) and is likely to be present in the acetone fraction. Six of the PSI core proteins (PsaD, PsaE, and PsaH) are found in pairs of closely related paralogues. Functional analysis has been performed for several of them (Varotto et al., 2000, 2002; Scheller et al., 2001; Haldrup et al., 2003). Despite the very high level of sequence identity between psaE1 and psaE2 and between psaH1 and psaH2 (89 and 99%, respectively), we found unique sequence tags (by ESI-MS/MS) for each and a unique (diagnostic) tag for psaD-2 (98% sequence identity with psaD-1) in the pepsin digest of fraction C/M 1:1 RP-HPLC.

All nine proteins of the ATP synthase complex were identified, including the very hydrophobic CFo-III and -IV subunits (Figure 7B). Four out of the eight subunits of the cytochrome *b6f* complex were found (Figure 7B). The four that were missing are all very small and hydrophobic.

As for the lumenal electron carriers, two plastocyanin (PC) proteins (but not cytochrome *c6*) were identified. Reverse genetic studies showed that the double mutant in both PC genes cannot grow autotrophically (Weigel et al., 2003). In *C. reinhardtii*, cytochrome *c6* can functionally complement PC under copper limited conditions (Quinn and Merchant, 1998). In *A. thaliana*, contrasting reports concerning such functional substitution of both PC proteins by cytochrome *c6* were reported (Gupta et al., 2002; Weigel et al., 2003). Two out of three FNR proteins were identified in the organic solvent fractions, in agreement with observations that chloroplasts expressed two FNR isoforms, whereas the third one is expressed in nonphotosynthetic plastids (Onda et al., 2000). An FNR protein was identified as a component of the cytochrome *b6f* complex in *Spinacia oleracea* (spinach) (Zhang et al., 2001) and is possibly the ortholog of At1g20020 (FNR-2). It is possible that one of the FNR paralogues associates with PSI and the other with the cytochrome *b6f* complex.

We can conclude that the coverage of the thylakoid bound photosynthetic apparatus was excellent. Using this experimental approach, it might now be possible to evaluate the role of different biogenesis components (targeting components, isomerases, protein translocators, etc.) and environmental conditions on the composition of these electron transport chain components.

Strikingly, when compared with all identified photosynthetic proteins, most of the missing proteins are small with high GRAVY values (Figure 6C), creating what appears to be an experimental bias against the very small proteins. The most likely explanation is that with tryptic digestion (C-terminal cleavage of the basic residues K and R), the fragments generated are typically too big to be detected by ESI-MS/MS. Indeed, many of the small hydrophobic proteins were identified after cleavage by other means (pepsin in the C/M 1:1 RP-HPLC fraction or acid hydrolysis/CnBr in the acetone fraction).

Mitochondrial Proteins and Other (Potentially) Nonchloroplast Proteins or Outer Envelope Proteins

Six mitochondrial proteins were found in our proteome analysis. Five of these are well-known abundant proteins: (1) the H subunit of the Gly decarboxylase complex (At2g35370) is one of the most abundant proteins of the plant mitochondria matrix (Oliver and Raman, 1995; Srinivasan and Oliver, 1995); (2) formate dehydrogenase (At5g14780), also identified in a 200-kD complex in *A. thaliana* mitochondria (Jansch et al., 1996); (3) subunit 6b of cytochrome c oxidase (At1g22450) in the inner membrane of plant mitochondria (Ohtsu et al., 2001); (4) subunit Vb of cytochrome c oxidase; and (5) a mitochondrial ATPase δ subunit (At3g52300), which also was identified (under accession number AL132972, a BAC clone) in a mitochondrial proteome analysis (Kruft et al., 2001). In addition, a member of the DegP protease family, DegP11, was identified. DegP11 is predicted to be localized in the mitochondria, but this is not proven experimentally.

Eleven proteins were identified that have no predicted cTP or mTP (prediction by TargetP) and they are not known chloroplast proteins or closely related paralogues. Five of those were found in an earlier study (Peltier et al., 2002) and are not discussed here. Two proteins (At3g63060 and At3g52230), found in one of the organic solvent fractions, are without any obvious function and might be outer envelope proteins because they also were detected in all three independent envelope proteome studies (Ferro et al., 2002, 2003; Froehlich et al., 2003). Such outer envelope proteins don't have recognizable cleavable cTPs. One protein was detected in the Na_2CO_3-extracted population and has a lipoxygenase domain (At4g39730) present in a variety of membrane- or lipid-associated proteins. A fatty acid elongase-like protein (At3g23840), an organic ion transporter-like protein (At3g20660), and a carbonic anhydrase (At4g20990) were identified in the RP-HPLC analysis of the C/M 1:1 fraction. From their predicted function, they might be chloroplast proteins, but further experimentation will be needed to determine their subcellular localization.

New Chloroplast Proteins

Twenty-six new proteins with a predicted cTP without known function were identified, with 24 of them exclusively identified in the organic solvent fractions (Table 1 in the supplemental data online). In addition, a new chloroplast protein (At3g47070) without predicted cTP was found. Seventeen of these 27 proteins are predicted to be membrane proteins and were only identified in the organic solvent fractions. These membrane proteins were predicted by TM-HMM to have between zero and three predicted TMs, whereas the consensus prediction in Aramemnon (http://aramemnon.botanik.uni-koeln.de/) predicted between one and three TMs (Table 1 in the supplemental data online). Three of the soluble proteins (At5g14910, At5g42765, and At2g34860) and one membrane protein (At4g22890) have predicted lTPs. Several proteins were identified in more than one organic solvent fraction, suggesting that they are real thylakoid components and not a low-level contaminant. None of them were identified by Schubert et al. (2002), but two proteins (At1g52220 and At1g42960) also were identified by Joyard and colleagues (Ferro et al., 2002, 2003), whereas eight others also were identified by Froehlich et al. (2003). One protein (At3g47070) without predicted cTP, identified both on the Na_2CO_3 2-DE map and in two of the organic solvent

fractions, was identified recently as a peripheral bound thylakoid phosphoprotein (Carlberg et al., 2003). At4g01150 and At2g46820 were identified in four of the organic solvent fractions and they are likely to be more abundant than the other new proteins. These four also were identified by Froehlich et al. (2003), possibly because they were relatively abundant or because they are both present in the inner envelopes and thylakoid.

A systematic search for functional domains using Pfam-HMM was performed and domains are listed in Table 1 in the supplemental data online. Seven of them (At1g54500, At5g17170, At5g36120, At4g01050, At5g14910, At2g34860, and At1g71500) have recognizable functional domains. The rubredoxin protein (At1g54500), also found by Froehlich et al. (2003), contains one predicted TM and is a homolog of a rubredoxin in *Synechococcus* sp PCC 7002, which was shown to be involved in the assembly of the interpolypeptide (4Fe-4S) cluster Fx of PSI (Shen et al., 2002a, 2002b). It is quite likely that it has the same function in thylakoids. At5g17170 has one predicted TM and also has a rubredoxin domain and a PDZ domain. The function of the protein is unknown but it is likely to be involved in assembly. At5g36120 is a protein with a predicted mature size of 135 amino acids, with two predicted TMs and a YGGT domain, and is a homolog of the *Synechocystis* protein S76575 with unknown function. The function of the YGGT domain is not known. At4g01050 has a rhodanese domain and is rich in Pro. TM-HMM predicts no TMs, whereas the consensus prediction suggest two TMs. Rhodaneses catalyze in vitro the transfer of a sulfur atom from certain sulfur donors to nucleophilic acceptors; *A. thaliana* has at least two rhodaneses, AtRDH1 and AtRDH2 (Hatzfeld and Saito, 2000). The function of At4g01050 is not clear. At1g42960 without known function also was identified by Ferro et al. (2002).

As for the soluble proteins, functional domains could be determined in 3 out of 10 proteins. One protein, At5g14910, has a predicted metal binding domain (HMA), a predicted lTP without a TAT signal, and also is identified in an ~140-kD complex on native gels (J.B. Peltier and K.J. van Wijk, unpublished data). This HMA domain is found in metallochaperones such as Atx1. Atx1 was proposed to support metal ion transfer between metal binding sites of soluble copper donors and membrane-tethered, copper-acceptor proteins (Portnoy et al., 1999; Rosenzweig et al., 1999). However, the HMA domain in At5g14910 does have several Cys but lacks the typical metal binding motif M/HCXXC (Wernimont et al., 2000). Just recently, a chloroplast copper binding protein, CUTA1, was reported that also lacks this HCXXC domain. Interestingly, CUTA1 shows some weak similarity to At5g14910 (data not shown). Other soluble proteins include a protein with a DnaJ domain (At2g34860) and a predicted lTP. There are 89 reported nuclear *A. thaliana* genes encoding for DnaJ-like proteins, with five having a predicted cTP (Miernyk, 2001). The DnaJ domain protein identified in our study (At2g34860) was not among these 89 proteins, most likely because annotation was less complete at the time of that analysis. DnaJ domain proteins interact with other chaperones, such as HSP70. Finally, one protein, also found by Froehlich et al. (2003), has some similarity to the Rieske protein (petC) of the cytochrome *b6f* complex (At1g71500) and the Rieske domain protein (At2g24820) found in the envelope (Caliebe et al., 1997).

Translation

This class contains 16 water-soluble proteins, nearly all identified on the 2-DE gels. Four of them also were found in the envelope preparation (Froehlich et al., 2003). These are comprised of ribosomal proteins, (putative) RNA binding proteins, three of the six known PSRPs (PSRP-2, PSRP-3, and PSRP-5), as well as a ribosome recycling factor. Ribosomal protein L12A/C was relatively abundant on the 2-DE gels and was found in several of the organic solvent fractions; it was the only ribosomal protein in these fractions. The protein has quite a high GRAVY index (+0.07), which explains why it does not precipitate in this apolar environment. L12 is located on the stalk of the 50S subunit and is central in the translocation step of translation. The most likely explanation for the presence of L12 in the organic solvent fraction is that L12 is present in four copies per ribosome (Yamaguchi and Subramanian, 2000). PSRP-2 was found in the lumenal enriched fraction as well as in the Na_2CO_3 and Triton population, suggesting a firm interaction with the thylakoid membrane. PSRP-3 only was found in the Na_2CO_3 map, whereas PSRP-5 only was found in the Triton fraction. The functions of these PSRPs are not clear but are likely related to stabilization of transcripts, regulation of translation, and possibly RNA processing (Yamaguchi and Subramanian, 2003). They all have RNA binding domains and PSRP-2 also has a poly(A) binding domain, similar to RB47 in *C. reinhardtii*. Their presence on the thylakoid surface also could suggest that they direct mRNA to the thylakoid, similar to some of the translation activators attached to the mitochondrial inner membrane in yeast (Naithani et al., 2003). The function of these PSRPs is currently under investigation.

The high percentage of proteins we identified to be involved in translation is consistent with the idea that the thylakoid membrane serves as a site for protein synthesis.

Primary and Secondary Metabolism

Only very few proteins involved in metabolism were identified in the thylakoid proteome, suggesting that the thylakoid membrane does not play a central role in primary and secondary metabolism. Metabolism-related proteins in the thylakoid fraction include carbonic anhydrase, both subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase and its activase, a lipoxygenase (AtLox2), and enoyl-acyl carrier protein reductase. These also were found in envelope articles (Ferro et al., 2002, 2003; Froehlich et al., 2003), probably because they are so abundant. AtLox2 previously was reported as a chloroplast protein, required for the wound-induced synthesis of the plant growth regulator jasmonic acid in leaves (Bell et al., 1995). We also find AtLox2 in relatively large quantities in the chloroplast stroma, mostly as monomers (J.-B. Peltier and K.J. van Wijk, unpublished data). Enoyl-acyl carrier reductase (At2g05990) is involved in fatty acid synthesis and its functional inactivation leads to premature cell death and dramatic alterations in plant morphology (Mou et al., 2000).

Chaperones, Protein Assembly, Insertion, Proteolysis, and Isomerases

This important class of proteins consists of 20 proteins with roles in protein folding, insertion, assembly, or proteolysis. All of the isomerases are located on the lumenal side of the thylakoid membrane and they might have a variety of functions, from folding of proteins to signaling (Fulgosi et al., 1998; Edvardsson et al., 2003). cpSecE is a component of the SecY/E translocon involved in the translocation of a subset of lumenal proteins as well as integral thylakoid membrane proteins (Schuenemann et al., 1999). Three FtsH proteases (FtsH2, FtsH5, and FtsH8) were tentatively identified. FtsH1 is involved in degradation of thylakoid proteins (Lindahl et al., 2000), and inactivation of FtsH2 (var 2) leads to a variegated phenotype (Chen et al., 2000). DegP1 and DegP5 are lumenal proteases (Itzhaki et al., 1998; Peltier et al., 2002; Schubert et al., 2002). ClpS1 is a protein tightly associated with the chloroplast ClpP/R complex and might function as a regulator of Clp proteolysis (Peltier et al., 2001, 2004). Chaperones HSP70 and Cpn60 have general protein chaperone functions and also are involved in protein import (Jackson-Constan et al., 2001). HSP70 and GrpE have been shown to form complexes at the thylakoid surface in *C. reinhardtii* in an ATP-dependent manner (Schroda et al., 2001). Finally, the rather abundant protein HCF136 functions in assembly of PSII (Meurer et al., 1998). Other potential assembly factors are two rubredoxins with one predicted TM and are discussed in the class of proteins with unknown function.

Alternative Electron Flow

Thylakoids have multiple strategies to avoid over-reduction of the electron transport chain and to adjust the production ratio of ATP:NADH. Three proteins involved in alternative electron flow were identified: PGR5 (At2g05620) and two of the chloroplast-encoded subunits of the NADPH-dehydrogenase (NDH) complex. Functional analysis of PGR5 recently showed that it is involved in the transfer of electrons from ferredoxin to plastoquinone to function in cyclic electron flow around PSI (Munekage et al., 2002). The NDH proteins are part of a 550-kD thylakoid-bound NDH complex that is involved in chlororespiration (Burrows et al., 1998; Peltier and Cournac, 2002). The NDH complex is dispensable for plant growth under optimal growth conditions (Burrows et al., 1998). Protein gel blot analysis suggested that the protein level of NDH-I was ~1.5% of the level of PSII (Burrows et al., 1998). Identification of these three proteins is encouraging because it shows that the dynamic resolution of this analysis is good.

Oxidative Stress

Seventeen proteins involved in stress avoidance (oxidative and light) were identified. Two of these, Ohp1 and Ohp2 (for one-helix proteins), are thylakoid membrane proteins with a single TM, most likely associated with PSI, and are upregulated under high light conditions (Jansson et al., 2000; Andersson et al., 2003). Other stress-related proteins include the three m-type thioredoxins (m1, m2, and m4) bound to the stromal side of the thylakoid membrane as well as 2-Cys peroxyredoxin (Horling et al., 2003) and FeSOD. They were predominantly present in the Yeda press fraction (for a discussion, see Peltier et al., 2002) rather than the Na_2CO_3-extracted proteome. This shows that these proteins are only very weakly associated with the thylakoid membrane surface. We identified expression of eight fibrillins, many of which were present in the lumenal enriched fraction, the Na_2CO_3 fraction, the Triton X-114 fraction, as well as the residual fraction after organic solvent extraction (Table 1; Table 1 in the supplemental data online). They were not identified in any of the organic solvent fractions, showing that the proteins are not very hydrophobic themselves (GRAVY values between −0.03 and −0.037). The fibrillins are present in electron-dense particles in so-called plastoglobules, which are free in the stroma as well as associated with the thylakoids and visible by TEM, but their function remains elusive (Rey et al., 2000).

Conclusion

This analysis gives a very good overview of the 150 to 200 most abundant proteins that serve as functional components of the thylakoid membrane system. Many unknown thylakoid proteins were identified. The challenge ahead now is to identify the function of these new proteins as well as the many paralogues of known photosynthetic proteins. We hope that the PPDB will become an important public resource for plastid protein data. It also is our intent to incorporate experimental identifications of plastid proteins in *A. thaliana* by other laboratories whenever they become available.

METHODS

Plant Growth and Thylakoid Preparations

A. thaliana (Columbia 0) was grown on soil with 10-h-light/14-h-dark periods at 22/19°C (light/dark). Thylakoids were purified from fully developed leaves using a combination of differential and gradient centrifugation steps according to Peltier et al. (2002), but with the following modifications to further improve thylakoid purity: centrifugation speed was reduced to 1850*g* for 3 min, and Percoll step cushions (20-40-80%) were used instead of linear Percoll gradients.

Extraction of Peripheral Thylakoid Proteomes and 2-DE Analysis

The lumenal-enriched thylakoid proteome was removed by Yeda press to remove the loosely attached stromal proteins as well as the majority of the soluble lumen proteins (Peltier et al., 2002). The remaining thylakoid membranes were resuspended at 1 mg of chlorophyll/mL in ice-cold 0.1 M Na_2CO_3 supplemented with a cocktail of nine protease inhibitors, as listed by Peltier et al. (2002). The suspension was stirred slowly for 30 min on ice with 15-s sonication steps after 0, 15, and 30 min of stirring. Membranes were removed by centrifugation at 150,000*g* for 25 min at 4°C. The supernatant was concentrated and desalted in centrifugal filter devices (Centricon, 3-kD cutoff; Millipore, Bedford, MA), followed by a 10-min spin at 200,000*g* to remove residual membranes. The supernatant then was precipitated in 80% acetone at −20°C. Precipitated proteins were collected by a 10-min centrifugation at 15,000*g* and resuspended in IPG rehydration buffer for protein determination and 2-DE analysis.

The remaining salt-extracted membrane pellets were resuspended in 7.2 mL of ice-cold Tris-buffered saline with the protease inhibitor cocktail, followed by the addition of 1.8 mL of precondensed Triton X-114 (Coligan

et al., 1995) to a final concentration of 0.14 mg of chlorophyll/mL. The suspension was homogenized 30 times. The suspension then was spun at 10,000*g* for 5 min in a cold centrifuge to remove the insoluble fractions, and supernatants were transferred to new tubes, warmed to 37°C for 3 min, and spun for 3 min at 1000*g*. The upper phase was collected and then mixed with one-fifth volume of Triton X-114, heated for 3 min at 37°C, and spun again for 3 min at 1000*g* (Coligan et al., 1995). The final upper phase was collected and concentrated and then precipitated in 90% cold (−20°C) acetone. The precipitated proteins were collected by centrifugation and resuspended in IPG rehydration buffer for 2-DE analysis.

The average yields of the Na_2CO_3-extracted peripheral and the Triton X-114–extracted proteome (upper phase) were 344 (SD 86; $n = 13$) and 24.8 (SD 10.5; $n = 6$) μg of protein per milligram of chlorophyll, respectively.

The peripheral proteomes, resuspended in IPG rehydration buffer, were focused on 18-cm IPG strips (Amersham-Pharmacia Biotech, Uppsala, Sweden) with pI from 3 to 10, 4 to 7, or 6 to 11, principally as described (Peltier et al., 2000, 2002). After focusing, the IPG strips were reduced and alkylated, and proteins were run out on 20 × 20-cm Tricine-SDS-PAGE gels (12% acryl amide) (Schägger and von Jagow, 1987). Gels were fixed and proteins visualized by high-sensitivity silver staining (Rabilloud et al., 1994), by MS-compatible silver staining (Shevchenko et al., 1996), by Sypro Ruby (Molecular Probes, Eugene, OR), or by Coomassie Brilliant Blue R 250. Gel images were taken either with the Personal Densitometer II (Molecular Dynamics, Sunnyvale, CA) or with the FluorS (Bio-Rad, Hercules, CA). Spot annotation of 2-DE gels was performed with the HT analyzer program (Genomic Solutions, Ann Arbor, MI).

Extraction and Separation of the Hydrophobic Membrane Proteome

Thylakoid membrane proteins were extracted from the Na_2CO_3-stripped thylakoid membranes using organic solvents. Thylakoid membranes (containing 300 mg of chlorophyll) were used as starting material. The pellet containing mostly integral membrane proteins was washed several times in 5 mM $MgCl_2$ and 100 mM Hepes-KOH (pH 8.0) to remove residual Na_2CO_3. Membranes then were solubilized in Laemmli buffer and heated at 42°C for 5 min. The insolubilized material was removed by a 14,000*g* spin for 5 min. Proteins in the supernatant were precipitated in nine volumes of cold acetone for 30 min at −20°C and pelleted at 12,000*g* for 30 min. Acetone in the supernatant then was evaporated, and the remaining liquid was dialyzed (500,000 times dilution) for 48 h and then concentrated. Subsequently, chlorophyll was removed by C/M extraction (see below). The acetone-soluble proteins in the methanol fraction then were dried down for MS analysis.

The acetone-precipitated pellet was resuspended in Laemmli solubilization buffer and used for sequential C/M extractions. These extractions were roughly based on Seigneurin-Berny et al. (1999), with a few changes. Each extraction with one specific C/M ratio was performed twice, and the remaining pellet was resuspended in Laemmli buffer, homogenized, and sonicated before the next extraction. C/M ratios of 9:1 and 1:1 were used. The residual C/M 1:9 insoluble pellet was resuspended in Laemmli buffer. The extracted protein fractions and the residual pellet were separated by 1-DE Tricine-PAGE (9 to 15% acryl amide).

RP-HPLC and In-Solution Digestion

In parallel to the gel-based separation, the C/M 1:1 extracted proteins were separated by RP-HPLC on an analytical microbore C4 column (Vydac 150 mm, 300 Å, total column volume ~15 μL; Grave Vydac, Columbia, MD) with 0.1% trifluoroacetic acid in water (solvent A) and 0.1% trifluoroacetic acid in isopropanol/acetonitrile 1:3 (solvent B). The C/M 1:1 fraction was dried out, resuspended in C/M 1:1, spun at 14,000 rpm for 15 min, and 20 or 50 μL supernatant was used for each HPLC run. Proteins were eluted using the following step gradient profile: 30 min at 5% B, 10 min from 5 to 35% B, 90 min from 35 to 75% B, 20 min from 75 to 95% B, and then 20 min at 95% at a flow rate of 50 μL/min. Fractions were collected continuously every 5 min immediately after injection, and the absorbance of each fraction monitored at 280 nm. Each fraction was dried down, dissolved in 70% formic acid, and then diluted by adding 20 volumes of 1 mM HCl. Proteins were digested with pepsin for 2 h at 37°C with a substrate-to-enzyme ratio of ~1:40, and samples were analyzed by ESI-MS/MS.

MS Analysis

Protein spots or bands from 1- and 2-DE gels were excised manually or in an automated fashion with a spot picking robot (ProPic; Genomic Solutions). The spots then were washed and digested with modified trypsin (Promega, Madison, WI) (Shevchenko et al., 1996) or pepsin, and peptides extracted automatically using the ProGest robot (Genomic Solutions). The acetone-soluble proteins were digested with trypsin and chymotrypsin or cleaved by limited acid hydrolysis (Shevchenko et al., 2000) or CnBr (van Montfort et al., 2002). Fractions from off-line RP-HPLC were digested in-solution with pepsin (in 70% formic acid and 1 mM HCl), trypsin (in 8 M urea), or chymotrypsin (in 8 M urea). For tryptic and chymotryptic digestion in-solution, the protein fractions were dried down, dissolved in 8M urea, and then diluted with ammonium bicarbonate to 1.6 M urea and 80 mM ammonium bicarbonate as digestion buffer (Yu et al., 1997). A substrate:protease ratio of ~100:1 was used, and the reaction was performed for 12 h at 37°C. Peptides were applied to the MALDI target plates by the dried droplet method using α-cyano-4-hydrocynnamic acid as matrix. The mass spectra were obtained automatically by MALDI-TOF MS in reflectron mode (Voyager-DE-STR; PerSeptive Biosystems, Framingham, MA) and were annotated automatically using the program m/z, as part of Knexus (Field et al., 2002) and http://www.genomicsolutions.com/, and internally calibrated using tryptic peptides from autodigestion. The latest versions of the National Center for Biotechnology Information BLAST nonredundant database or Arabidopsis Munich Information Center for Protein Sequences or TIGR databases (downloaded locally) were searched automatically with the resulting peptide mass lists, using the search engine ProFound (Zhang and Chait, 2000), as part of Knexus (Field et al., 2002).

For further investigation, the peptides also were analyzed manually with nanoelectrospray needles (Protana, Odense, Denmark) or by nano-LC-ESI-MS/MS in automated mode on a quadruple/orthogonal-acceleration TOF tandem mass spectrometer (Q-TOF; Micromass, Manchester, UK). The spectra were used to search the public databases (downloaded locally) automated via Sonar MS/MS, as part of Knexus (Field et al., 2002), or manually via Mascot (www.matrixscience.com).

Criteria for positive identification by MALDI-TOF MS peptide mass fingerprinting are five or more matching peptides with a narrow error distribution (clustering of errors) within 25 ppm and at least 15% sequence coverage. In exceptional cases, four matching peptides were considered as positive identification (e.g., proteins <20 kD and matching gel coordinates), along with high sequence coverage. Only peptides that have no missed cleavages (by trypsin) and no modifications, other than Met oxidation and carbamidomethylation (because gel samples were alkylated), were considered. Criteria for positive identification by MS/MS were as follows: if only one or two matching sequence tags were found, nearly complete Y-ion series (in the case of trypsin digest) and partial complementary B-ion series needed to be present in each (as determined by manual inspection), with several diagnostic immonium ions present. Peptide scores in Mascot needed to be >35 and total protein scores (probability-based Mowse score) >37 (when searching only on all *A. thaliana* gene models), but we never relied on score alone. In the case of three or more matching peptides with peptide scores >35, manual

inspection of Y- and B-ion series was not required. Most of the product ions matched within <0.05 D, and precursor ions matched within 1 D. When scoring with Sonar MS/MS, protein scores need to be smaller than E^{-3}, and peptide scores needed to be well separated from the statistical noise (visualized in Sonar MS/MS as [Log]frequency being much lower than [Log]score). Again, peptide matches were manually verified.

For our internal database and in Tables 2 and 3 in the supplemental data online, all matches were translated in a confidence level 1 or 2, with level 1 being the most confident. Confidence level 2 was chosen if the identification was close to the minimum matching requirements (e.g., four to six matching peptides in MALDI-TOF) and level 1 was chosen if it was well over these minimal requirements (e.g., >7 matching peptides and high sequence coverage or two or more strong MS/MS sequence tags). Ambiguity number refers to matches to two or more closely related paralogues; these paralogues were linked with the same integer number if it was difficult to distinguish between the paralogues (when peptide sequence(s) or peptide masses were shared by two or more proteins). Examples are proteins in the LHC family.

Construction of the PPDB and Theoretical Analysis of the Proteins

The database engine for PPDB is a Microsoft SQL server. The Web interface for PPDB was developed on ASP.NET platform using C# language. Molecular weight, pI, GRAVY (Kyte and Doolittle, 1982), and other biophysical properties were calculated using the Emboss software suite (Rice et al., 2000). TargetP (Emanuelsson et al., 2000) was used for chloroplast protein prediction and cTP cleavage site prediction, and LumenP was used for predicting lumen proteins (Westerlund et al., 2003). Transmembrane domain information was based on predictions by TMHMM (Krogh et al., 2001). The functional domain prediction was based on Pfam analysis with a cutoff E-value of 0.1 (Bateman et al., 2002). The experimental data in PPDB (including curated gene information) are provided by members of the K.J. van Wijk lab. The experimental data are linked to the latest annotation of the *A. thaliana* genome (currently release 4.0 of ATH1.pep, with 28,581 nuclear-encoded A. thaliana proteins) provided by TIGR (http://www.tigr.org/tdb/e2k1/ath1/) as well as the mitochondrial and plastid genomes.

Miscellaneous

Chlorophyll concentrations were determined spectrometrically in 80% acetone (Porra et al., 1989), and protein determinations were performed with the Bradford assay (Bradford, 1976) or with bicinchoninic acid (Smith et al., 1985). Protein gel blot analysis with polyclonal antisera against Toc75 and OEC23 was performed as described earlier (Peltier et al., 2000).

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession number AL132972.

ACKNOWLEDGMENTS

Ted Tennhauser and Michael J. Mahar from the Biotechnology Resource Center at Cornell are gratefully acknowledged for assistance with off-line RP-HPLC. We thank Danny Schnell for his generous gift of antiserum against *A. thaliana* Toc75. This study was supported by the National Science Foundation Division of Molecular and Cellular Biosciences 0090942, Cornell University, and the New York State Office of Science, Technology and Research to K.J.V.W. All database research was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners.

Received September 28, 2003; accepted November 6, 2003.

REFERENCES

Andersson, U., Heddad, M., and Adamska, I. (2003). Light stress-induced one-helix protein of the chlorophyll a/b-binding family associated with photosystem I. Plant Physiol. **132,** 811–820.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408,** 796–815.

Aro, E.-M., and Andersson, B. (2001). Regulation of Photosynthesis. (Dordrecht, The Netherlands: Kluwer Academic Publishers).

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Res. **30,** 276–280.

Bell, E., Creelman, R.A., and Mullet, J.E. (1995). A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in Arabidopsis. Proc. Natl. Acad. Sci. USA **92,** 8675–8679.

Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72,** 248–254.

Burkhead, J.L., Abdel-Ghany, S.E., Morrill, J.M., Pilon-Smits, E.A., and Pilon, M. (2003). The Arabidopsis thaliana CUTA gene encodes an evolutionarily conserved copper binding chloroplast protein. Plant J. **34,** 856–867.

Burrows, P.A., Sazanov, L.A., Svab, Z., Maliga, P., and Nixon, P.J. (1998). Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid ndh genes. EMBO J. **17,** 868–876.

Caliebe, A., Grimm, R., Kaiser, G., Lubeck, J., Soll, J., and Heins, L. (1997). The chloroplastic protein import machinery contains a Rieske-type iron-sulfur cluster and a mononuclear iron-binding protein. EMBO J. **16,** 7342–7350.

Carlberg, I., Hansson, M., Kieselbach, T., Schroder, W.P., Andersson, B., and Vener, A.V. (2003). A novel plant protein undergoing light-induced phosphorylation and release from the photosynthetic thylakoid membranes. Proc. Natl. Acad. Sci. USA **100,** 757–762.

Chen, M., Choi, Y., Voytas, D.F., and Rodermel, S. (2000). Mutations in the Arabidopsis VAR2 locus cause leaf variegation due to the loss of a chloroplast FtsH protease. Plant J. **22,** 303–313.

Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., and Wingfield, P.T. (1995). Current Protocols in Protein Science. (New York: John Wiley & Sons).

Depege, N., Bellafiore, S., and Rochaix, J.D. (2003). Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in Chlamydomonas. Science **299,** 1572–1575.

Edvardsson, A., Eshaghi, S., Vener, A.V., and Andersson, B. (2003). The major peptidyl-prolyl isomerase activity in thylakoid lumen of plant chloroplasts belongs to a novel cyclophilin TLP20. FEBS Lett. **542,** 137–141.

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

Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., and Rolland, N. (2003). Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. Mol Cell Proteomics **2,** 325–345.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002). Integral membrane proteins of the chloroplast envelope:

Identification and subcellular localization of new transporters. Proc. Natl. Acad. Sci. USA **99,** 11487–11492.

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J., and Joyard, J. (2000). Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis **21,** 3517–3526.

Field, H.I., Fenyo, D., and Beavis, R.C. (2002). RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. Proteomics **2,** 36–47.

Fischer, K., and Weber, A. (2002). Transport of carbon in non-green plastids. Trends Plant Sci. **7,** 345–351.

Flugge, U.I. (1998). Metabolite transporters in plastids. Curr. Opin. Plant Biol. **1,** 201–206.

Froehlich, J.E., Wilkerson, C.G., Ray, W.K., McAndrew, R.S., Osteryoung, K.W., Gage, D.A., and Phinney, B.S. (2003). Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. J. Proteome Res. **2,** 413–425.

Fulgosi, H., Vener, A.V., Altschmied, L., Herrmann, R.G., and Andersson, B. (1998). A novel multi-functional chloroplast protein: Identification of a 40 kDa immunophilin-like protein located in the thylakoid lumen. EMBO J. **17,** 1577–1587.

Gomez, S.M., Nishio, J.N., Faull, K.F., and Whitelegge, J.P. (2002). The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. Mol. Cell Proteomics **1,** 46–59.

Gupta, R., He, Z., and Luan, S. (2002). Functional relationship of cytochrome c(6) and plastocyanin in Arabidopsis. Nature **417,** 567–571.

Haldrup, A., Lunde, C., and Scheller, H.V. (2003). Arabidopsis thaliana plants lacking the PSI-D subunit of photosystem I suffer severe photoinhibition, have unstable photosystem I complexes, and altered redox homeostasis in the chloroplast stroma. J. Biol. Chem. **278,** 33276–33283.

Hankamer, B., Morris, E., Nield, J., Carne, A., and Barber, J. (2001). Subunit positioning and transmembrane helix organisation in the core dimer of photosystem II. FEBS Lett. **504,** 142–151.

Hatzfeld, Y., and Saito, K. (2000). Evidence for the existence of rhodanese (thiosulfate:cyanide sulfurtransferase) in plants: Preliminary characterization of two rhodanese cDNAs from Arabidopsis thaliana. FEBS Lett. **470,** 147–150.

Horling, F., Lamkemeyer, P., Konig, J., Finkemeier, I., Kandlbinder, A., Baier, M., and Dietz, K.J. (2003). Divergent light-, ascorbate-, and oxidative stress-dependent regulation of expression of the peroxiredoxin gene family in Arabidopsis. Plant Physiol. **131,** 317–325.

Issakidis-Bourguet, E., Mouaheb, N., Meyer, Y., and Miginiac-Maslow, M. (2001). Heterologous complementation of yeast reveals a new putative function for chloroplast m-type thioredoxin. Plant J. **25,** 127–135.

Itzhaki, H., Naveh, L., Lindahl, M., Cook, M., and Adam, Z. (1998). Identification and characterization of DegP, a serine protease associated with the luminal side of the thylakoid membrane. J. Biol. Chem. **273,** 7094–7098.

Jackson-Constan, D., Akita, M., and Keegstra, K. (2001). Molecular chaperones involved in chloroplast protein import. Biochim. Biophys. Acta **1541,** 102–113.

Jackson-Constan, D., and Keegstra, K. (2001). Arabidopsis genes encoding components of the chloroplastic protein import apparatus. Plant Physiol. **125,** 1567–1576.

Jansch, L., Kruft, V., Schmitz, U.K., and Braun, H.P. (1996). New insights into the composition, molecular mass and stoichiometry of the protein complexes of plant mitochondria. Plant J. **9,** 357–368.

Jansson, S., Andersson, J., Kim, S.J., and Jackowski, G. (2000). An Arabidopsis thaliana protein homologous to cyanobacterial high-light-inducible proteins. Plant Mol. Biol. **42,** 345–351.

Jarvis, P., and Soll, J. (2002). Toc, tic, and chloroplast protein import. Biochim. Biophys. Acta **1590,** 177–189.

Kamiya, N., and Shen, J.R. (2003). Crystal structure of oxygen-evolving photosystem II from Thermosynechococcus vulcanus at 3.7-A resolution. Proc. Natl. Acad. Sci. USA **100,** 98–103.

Karpinski, S., Reynolds, H., Karpinska, B., Wingsle, G., Creissen, G., and Mullineaux, P. (1999). Systemic signaling and acclimation in response to excess excitation energy in Arabidopsis. Science **284,** 654–657.

Knoetzel, J., Mant, A., Haldrup, A., Jensen, P.E., and Scheller, H.V. (2002). PSI-O, a new 10-kDa subunit of eukaryotic photosystem I. FEBS Lett. **510,** 145–148.

Konig, J., Baier, M., Horling, F., Kahmann, U., Harris, G., Schurmann, P., and Dietz, K.J. (2002). The plant-specific function of 2-Cys peroxiredoxin-mediated detoxification of peroxides in the redox-hierarchy of photosynthetic electron flux. Proc. Natl. Acad. Sci. USA **99,** 5738–5743.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. **305,** 567–580.

Kruft, V., Eubel, H., Jansch, L., Werhahn, W., and Braun, H.P. (2001). Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. Plant Physiol. **127,** 1694–1710.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157,** 105–132.

Leister, D. (2003). Chloroplast research in the genomic age. Trends Genet. **19,** 47–56.

Lennartz, K., Plucken, H., Seidler, A., Westhoff, P., Bechtold, N., and Meierhoff, K. (2001). HCF164 encodes a thioredoxin-like protein involved in the biogenesis of the cytochrome b(6)f complex in Arabidopsis. Plant Cell **13,** 2539–2551.

Lew, S., and London, E. (1997). Simple procedure for reversed-phase high-performance liquid chromatographic purification of long hydrophobic peptides that form transmembrane helices. Anal. Biochem. **251,** 113–116.

Lindahl, M., Spetea, C., Hundal, T., Oppenheim, A.B., Adam, Z., and Andersson, B. (2000). The thylakoid FtsH protease plays a role in the light-induced turnover of the photosystem II D1 protein. Plant Cell **12,** 419–432.

Meurer, J., Plucken, H., Kowallik, K.V., and Westhoff, P. (1998). A nuclear-encoded protein of prokaryotic origin is essential for the stability of photosystem II in Arabidopsis thaliana. EMBO J. **17,** 5286–5297.

Miernyk, J.A. (2001). The J-domain proteins of Arabidopsis thaliana: An unexpectedly large and diverse family of chaperones. Cell Stress Chaperones **6,** 209–218.

Millar, A.H., Sweetlove, L.J., Giege, P., and Leaver, C.J. (2001). Analysis of the Arabidopsis mitochondrial proteome. Plant Physiol. **127,** 1711–1727.

Molloy, M.P., Herbert, B.R., Williams, K.L., and Gooley, A.A. (1999). Extraction of Escherichia coli proteins with organic solvents prior to two-dimensional electrophoresis. Electrophoresis **20,** 701–704.

Monde, R.A., Schuster, G., and Stern, D.B. (2000). Processing and degradation of chloroplast mRNA. Biochimie **82,** 573–582.

Mori, H., and Cline, K. (2001). Post-translational protein translocation into thylakoids by the Sec and DeltapH-dependent pathways. Biochim. Biophys. Acta **1541,** 80–90.

Mou, Z., He, Y., Dai, Y., Liu, X., and Li, J. (2000). Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. Plant Cell **12,** 405–418.

Munekage, Y., Hojo, M., Meurer, J., Endo, T., Tasaka, M., and Shikanai, T. (2002). PGR5 is involved in cyclic electron flow around photosystem I and is essential for photoprotection in Arabidopsis. Cell **110,** 361–371.

Naithani, S., Saracco, S.A., Butler, C.A., and Fox, T.D. (2003). Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of Saccharomyces cerevisiae. Mol. Biol. Cell **14,** 324–333.

Nield, J., Orlova, E.V., Morris, E.P., Gowen, B., van Heel, M., and Barber, J. (2000). 3D map of the plant photosystem II supercomplex obtained by cryoelectron microscopy and single particle analysis. Nat. Struct. Biol. **7,** 44–47.

Ohtsu, K., Nakazono, M., Tsutsumi, N., and Hirai, A. (2001). Characterization and expression of the genes for cytochrome c oxidase subunit VIb (COX6b) from rice and Arabidopsis thaliana. Gene **264,** 233–239.

Oliver, D.J., and Raman, R. (1995). Glycine decarboxylase: Protein chemistry and molecular biology of the major protein in leaf mitochondria. J. Bioenerg. Biomembr. **27,** 407–414.

Onda, Y., Matsumura, T., Kimata-Ariga, Y., Sakakibara, H., Sugiyama, T., and Hase, T. (2000). Differential interaction of maize root ferredoxin:NADP(+) oxidoreductase with photosynthetic and non-photosynthetic ferredoxin isoproteins. Plant Physiol. **123,** 1037–1045.

Ort, D.R., and Yocum, C.F. 1996. Oxygenic Photosynthesis: The Light Reactions. (Dordrecht, The Netherlands: Kluwer Academic Publishers).

Peltier, G., and Cournac, L. (2002). Chlororespiration. Annu. Rev. Plant Biol. **53,** 523–550.

Peltier, J.B., Emanuelsson, O., Kalume, D.E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D.A., Soderberg, L., Roepstorff, P., von Heijne, G., and van Wijk, K.J. (2002). Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. Plant Cell **14,** 211–236.

Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K.J. (2000). Proteomics of the chloroplast: Systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell **12,** 319–342.

Peltier, J.B., Ripoll, D.R., Friso, G., Rudella, A., Cai, Y., Ytterberg, A.J., Giacomelli, L., Pillardy, J., and van Wijk, K.J. (2004). Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted 3-D structures and functional implications. J. Biol. Chem., in press.

Peltier, J.B., Ytterberg, J., Liberles, D.A., Roepstorff, P., and van Wijk, K.J. (2001). Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of Arabidopsis thaliana. J. Biol. Chem. **276,** 16318–16327.

Porra, R.J., Thompson, W.A., and Kriedemann, P.E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls *a* and *b* extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. Biochim. Biophys. Acta **975,** 384–394.

Portnoy, M.E., Rosenzweig, A.C., Rae, T., Huffman, D.L., O'Halloran, T.V., and Culotta, V.C. (1999). Structure-function analyses of the ATX1 metallochaperone. J. Biol. Chem. **274,** 15041–15045.

Quinn, J.M., and Merchant, S. (1998). Copper-responsive gene expression during adaptation to copper deficiency. Methods Enzymol. **297,** 263–279.

Rabilloud, T., Vuillard, L., Gilly, C., and Lawrence, J.J. (1994). Silver-staining of proteins in polyacrylamide gels: A general overview. Cell Mol. Biol. (Noisy-le-grand) **40,** 57–75.

Rey, P., Gillet, B., Romer, S., Eymery, F., Massimino, J., Peltier, G., and Kuntz, M. (2000). Over-expression of a pepper plastid lipid-associated protein in tobacco leads to changes in plastid ultrastructure and plant development upon stress. Plant J. **21,** 483–494.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. **16,** 276–277.

Rochaix, J.-D., Goldschmidt-Clermont, M., and Merchant, S. (1998). The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas, Advances in Photosynthesis and Respiration, Vol. 7. (Dordrecht, The Netherlands: Kluwer Academic Publishers).

Rosenzweig, A.C., Huffman, D.L., Hou, M.Y., Wernimont, A.K., Pufahl, R.A., and O'Halloran, T.V. (1999). Crystal structure of the Atx1 metallochaperone protein at 1.02 A resolution. Structure Fold. Des. **7,** 605–617.

Sakamoto, W., Tamura, T., Hanba-Tomita, Y., and Murata, M. (2002). The VAR1 locus of Arabidopsis encodes a chloroplastic FtsH and is responsible for leaf variegation in the mutant alleles. Genes Cells **7,** 769–780.

Schägger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. **166,** 368–379.

Scheller, H.V., Jensen, P.E., Haldrup, A., Lunde, C., and Knoetzel, J. (2001). Role of subunits in eukaryotic Photosystem I. Biochim. Biophys. Acta **1507,** 41–60.

Schneider, A., Hausler, R.E., Kolukisaoglu, U., Kunze, R., van der Graaff, E., Schwacke, R., Catoni, E., Desimone, M., and Flugge, U.I. (2002). An Arabidopsis thaliana knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. Plant J. **32,** 685–699.

Schroda, M., Vallon, O., Whitelegge, J.P., Beck, C.F., and Wollman, F.A. (2001). The chloroplastic GrpE homolog of Chlamydomonas: Two isoforms generated by differential splicing. Plant Cell **13,** 2823–2839.

Schubert, M., Petersson, U.A., Haas, B.J., Funk, C., Schroder, W.P., and Kieselbach, T. (2002). Proteome map of the chloroplast lumen of Arabidopsis thaliana. J. Biol. Chem. **277,** 8354–8365.

Schuenemann, D., Amin, P., Hartmann, E., and Hoffman, N.E. (1999). Chloroplast SecY is complexed to SecE and involved in the translocation of the 33-kDa but not the 23-kDa subunit of the oxygen-evolving complex. J. Biol. Chem. **274,** 12177–12182.

Seigneurin-Berny, D., Rolland, N., Garin, J., and Joyard, J. (1999). Technical Advance: Differential extraction of hydrophobic proteins from chloroplast envelope membranes: A subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. Plant J. **19,** 217–228.

Shen, G., Antonkine, M.L., van der Est, A., Vassiliev, I.R., Brettel, K., Bittl, R., Zech, S.G., Zhao, J., Stehlik, D., Bryant, D.A., and Golbeck, J.H. (2002a). Assembly of photosystem I. II. Rubredoxin is required for the in vivo assembly of F(X) in Synechococcus sp. PCC 7002 as shown by optical and EPR spectroscopy. J. Biol. Chem. **277,** 20355–20366.

Shen, G., Zhao, J., Reimer, S.K., Antonkine, M.L., Cai, Q., Weiland, S.M., Golbeck, J.H., and Bryant, D.A. (2002b). Assembly of photosystem I. I. Inactivation of the rubA gene encoding a membrane-associated rubredoxin in the cyanobacterium Synechococcus sp. PCC 7002 causes a loss of photosystem I activity. J. Biol. Chem. **277,** 20343–20354.

Shevchenko, A., Loboda, A., Ens, W., and Standing, K.G. (2000). MALDI quadrupole time-of-flight mass spectrometry: A powerful tool for proteomic research. Anal. Chem. **72,** 2132–2141.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. **68,** 850–858.

Shikanai, T., Muller-Moule, P., Munekage, Y., Niyogi, K.K., and Pilon, M. (2003). PAA1, a P-type ATPase of Arabidopsis, functions in copper transport in chloroplasts. Plant Cell **15,** 1333–1346.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985). Measurement of protein using bicinchoninic acid. Anal. Biochem. **150,** 76–85.

Snyders, S., and Kohorn, B.D. (1999). TAKs, thylakoid membrane protein kinases associated with energy transduction. J. Biol. Chem. **274,** 9137–9140.

Srinivasan, R., and Oliver, D.J. (1995). Light-dependent and tissue-specific expression of the H-protein of the glycine decarboxylase complex. Plant Physiol. **109,** 161–168.

Sussman, M.R. (1988). Purification of integral plasma membrane proteins by reverse-phase high performance liquid chromatography. Anal. Biochem. **169,** 395–399.

Swiatek, M., Kuras, R., Sokolenko, A., Higgs, D., Olive, J., Cinque, G., Muller, B., Eichacker, L.A., Stern, D.B., Bassi, R., Herrmann, R.G., and Wollman, F.A. (2001). The chloroplast gene ycf9 encodes a photosystem II (PSII) core subunit, PsbZ, that participates in PSII supramolecular architecture. Plant Cell **13,** 1347–1367.

Tarr, G.E., and Crabb, J.W. (1983). Reverse-phase high-performance liquid chromatography of hydrophobic proteins and fragments thereof. Anal. Biochem. **131,** 99–107.

Teramoto, H., Nakamori, A., Minagawa, J., and Ono, T.A. (2002). Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in Chlamydomonas reinhardtii. Plant Physiol. **130,** 325–333.

Tyers, M., and Mann, M. (2003). From genomics to proteomics. Nature **422,** 193–197.

van Montfort, B.A., Canas, B., Duurkens, R., Godovac-Zimmermann, J., and Robillard, G.T. (2002). Improved in-gel approaches to generate peptide maps of integral membrane proteins with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J. Mass Spectrom. **37,** 322–330.

van Wijk, K.J. (2001a). Proteins involved in biogenesis of the thylakoid membrane. In Regulatory Aspects of Photosynthesis, Advances in Photosynthesis and Respiration, Vol. 11. E.-M. Aro and B. Andersson, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 153–175.

van Wijk, K.J. (2001b). Challenges and prospects of plant proteomics. Plant Physiol. **126,** 501–508.

Varotto, C., Pesaresi, P., Jahns, P., Lessnick, A., Tizzano, M., Schiavon, F., Salamini, F., and Leister, D. (2002). Single and double knockouts of the genes for photosystem I subunits G, K, and H of Arabidopsis. Effects on photosystem I composition, photosynthetic electron flow, and state transitions. Plant Physiol. **129,** 616–624.

Varotto, C., Pesaresi, P., Meurer, J., Oelmuller, R., Steiner-Lange, S., Salamini, F., and Leister, D. (2000). Disruption of the Arabidopsis photosystem I gene psaE1 affects photosynthesis and impairs growth. Plant J. **22,** 115–124.

Watson, B.S., Asirvatham, V.S., Wang, L., and Sumner, L.W. (2003). Mapping the proteome of barrel medic (Medicago truncatula). Plant Physiol. **131,** 1104–1123.

Weigel, M., Varotto, C., Pesaresi, P., Finazzi, G., Rappaport, F., Salamini, F., and Leister, D. (2003). Plastocyanin is indispensable for photosynthetic electron flow in Arabidopsis thaliana. J. Biol. Chem. **278,** 31286–31289.

Wernimont, A.K., Huffman, D.L., Lamb, A.L., O'Halloran, T.V., and Rosenzweig, A.C. (2000). Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins. Nat. Struct. Biol. **7,** 766–771.

Westerlund, I., Von Heijne, G., and Emanuelsson, O. (2003). LumenP — a neural network predictor for protein localization in the thylakoid lumen. Protein Sci. **12,** 2360–2366.

Whitelegge, J.P., Zhang, H., Aguilera, R., Taylor, R.M., and Cramer, W.A. (2002). Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein: Cytochrome b(6)f complex from spinach and the cyanobacterium Mastigocladus laminosus. Mol. Cell Proteomics **1,** 816–827.

Wollman, F.A., Minai, L., and Nechushtai, R. (1999). The biogenesis and assembly of photosynthetic proteins in thylakoid membranes1. Biochim. Biophys. Acta **1411,** 21–85.

Yabuta, Y., Motoki, T., Yoshimura, K., Takeda, T., Ishikawa, T., and Shigeoka, S. (2002). Thylakoid membrane-bound ascorbate peroxidase is a limiting factor of antioxidative systems under photooxidative stress. Plant J. **32,** 915–925.

Yamaguchi, K., and Subramanian, A.R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. **275,** 28466–28482.

Yamaguchi, K., and Subramanian, A.R. (2003). Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30S ribosomal subunit. Eur. J. Biochem. **270,** 190–205.

Yu, Z., Friso, G., Miranda, J.J., Patel, M.J., Lo-Tseng, T., Moore, E.G., and Burlingame, A.L. (1997). Structural characterization of human hemoglobin crosslinked by bis(3,5- dibromosalicyl) fumarate using mass spectrometric techniques. Protein Sci. **6,** 2568–2577.

Zhang, H., Whitelegge, J.P., and Cramer, W.A. (2001). Ferredoxin: NADP+ oxidoreductase is a subunit of the chloroplast cytochrome b6f complex. J. Biol. Chem. **276,** 38159–38165.

Zhang, W., and Chait, B.T. (2000). ProFound: An expert system for protein identification using mass spectrometric peptide mapping information. Anal. Chem. **72,** 2482–2489.