

Deciphering Thylakoid Sub-compartments using a Mass Spectrometry-based Approach^{*}

Martino Tomizoli^{†§||}, Cosmin Lazar^{†**‡‡}, Sabine Brugière^{†**‡‡},
Thomas Burger^{†**‡‡§§}, Daniel Salvi^{†§||}, Laurent Gatto^{||}, Lucas Moyet^{†§||},
Lisa M. Breckels^{||}, Anne-Marie Hesse^{†**‡‡}, Kathryn S. Lilley^{||},
Daphné Seigneurin-Berny^{†§||}, Giovanni Finazzi^{†§||}, Norbert Rolland^{†§|||},
and Myriam Ferro^{†**‡‡||}

Photosynthesis has shaped atmospheric and ocean chemistries and probably changed the climate as well, as oxygen is released from water as part of the photosynthetic process. In photosynthetic eukaryotes, this process occurs in the chloroplast, an organelle containing the most abundant biological membrane, the thylakoids. The thylakoids of plants and some green algae are structurally inhomogeneous, consisting of two main domains: the grana, which are piles of membranes gathered by stacking forces, and the stroma-lamellae, which are unstacked thylakoids connecting the grana. The major photosynthetic complexes are unevenly distributed within these compartments because of steric and electrostatic constraints. Although proteomic analysis of thylakoids has been instrumental to define its protein components, no extensive proteomic study of subthylakoid localization of proteins in the BBY (grana) and the stroma-lamellae fractions has been achieved so far. To fill this gap, we performed a complete survey of the protein composition of these thylakoid subcompartments using thylakoid membrane fractionations. We employed semiquantitative proteomics coupled with a data analysis pipeline and manual annotation to differentiate genuine BBY and stroma-lamellae proteins from possible contaminants. About 300 thylakoid (or potentially thylakoid) proteins were shown to

be enriched in either the BBY or the stroma-lamellae fractions. Overall, present findings corroborate previous observations obtained for photosynthetic proteins that used nonproteomic approaches. The originality of the present proteomic relies in the identification of photosynthetic proteins whose differential distribution in the thylakoid subcompartments might explain already observed phenomenon such as LHCII docking. Besides, from the present localization results we can suggest new molecular actors for photosynthesis-linked activities. For instance, most PsbP-like subunits being differently localized in stroma-lamellae, these proteins could be linked to the PSI-NDH complex in the context of cyclic electron flow around PSI. In addition, we could identify about a hundred new likely minor thylakoid (or chloroplast) proteins, some of them being potential regulators of the chloroplast physiology. *Molecular & Cellular Proteomics* 13: 10.1074/mcp.M114.040923, 2147–2167, 2014.

As primary producers, plants are at the basis of the food chain for most ecosystems and control the planet atmosphere via photosynthesis. In eukaryotes, this process, responsible for carbon fixation and for the release of gaseous oxygen into the atmosphere, takes place in a specialized compartment, the chloroplast. This semiautonomous organelle plays a number of essential functions, including photosynthesis, nitrogen assimilation, sulfur reduction and assimilation, synthesis of amino acids, fatty acids, and many secondary metabolites (1–6). The chloroplast is not the only type of plastids. Plastids are indeed found in every plant tissue (with a very few exceptions such as angiosperm pollen grains) and in apicomplexan parasites (7, 8). Plastids play diverse and developmentally regulated functions including carotenoids accumulation in flowers and fruits (chromoplasts) and starch accumulation (amyloplasts). Chloroplasts are the most studied plastid type. They are distributed throughout the cytoplasm of leaf cells and contain several key subcompartments including: (1) the chloroplast envelope, which is a double membrane system surrounding the organelle and modulating the communication of the chloroplast with the plant cell; (2) the stroma, mainly

From the [†]Univ. Grenoble Alpes, F-38000 Grenoble, France; [§]CNRS, UMR5168, F-38054 Grenoble, France; ^{||}CEA, iRTSV, Laboratoire Physiologie Cellulaire & Végétale, F-38054 Grenoble, France; ^{||}INRA, USC 1359, F-38054 Grenoble, France; ^{**}CEA, iRTSV, Laboratoire Biologie à Grande Echelle, F-38054 Grenoble, France; ^{‡‡}INSERM, U1038, F-38054 Grenoble, France; ^{§§}CNRS, FR3425, F-38054 Grenoble, France; ^{||}Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, CB2 1QR, United Kingdom

Received, May 2, 2014 and in revised form, May 2, 2014

Published, MCP Papers in Press, May 28, 2014 DOI 10.1074/mcp.M114.040923

Author contributions: G.F., N.R., and M.F. designed research; M.T., S.B., D. Salvi, L.M., D. Seigneurin-Berny, and G.F. performed research; M.T., C.L., T.B., A.H., D. Seigneurin-Berny, G.F., N.R., and M.F. analyzed data; M.T., C.L., T.B., D. Seigneurin-Berny, G.F., N.R., and M.F. wrote the paper; L.G. contributed software development; L.M.B. and A.H. contributed to software development; K.S.L. supervision of software development.

composed of soluble proteins, which is the site where CO₂ assimilation takes place thanks to the consumption of reducing equivalents and ATP produced by light-driven electron flow, and (3) the thylakoid membrane, which is a highly organized internal membrane network formed of flat compressed vesicles and which is the center of oxygenic photosynthesis. The thylakoid vesicles delimit another discrete compartment, the lumen. Being the place of the light phase of photosynthesis, the thylakoid compartment has been deeply investigated from a functional, structural, and biochemical point of view. Several complexes are located in the thylakoid membranes to catalyze this activity: the two photosystems, PSI and PSII, the cytochrome *b₆f*, and the ATP synthase CF₀-F₁. They mostly act in series and their function must be tightly regulated to avoid excess (or unbalanced) light absorption or saturation of the electron flow chain, which leads otherwise to photodamage (9). To do so, the rather simple photosynthetic membrane present in cyanobacteria has progressively evolved into a more sophisticated structure in some algae and plants. There, the thylakoid membranes form a physically continuous three-dimensional network consisting of two main domains: the grana, which are stacks of thylakoids, and the stroma-lamellae, which are unstacked thylakoids connecting the grana stacks. The grana stacks are surrounded by “unstacked” membranes, called grana ends and margins (10).

High throughput analysis techniques, combined with genome sequencing data, have largely modified the experimental approaches to study protein localization within a cell and changed our perception of the functional consequence of this localization (11–13). This is also true for the chloroplast, where several proteome catalogs have been produced so far (14–17). They include descriptions of whole chloroplast proteomes, as well as of its subcompartments. In particular, thylakoids have been subjected to numerous proteomic investigations, mostly focusing on two biochemically different thylakoid protein populations: soluble proteins located in the lumen and more hydrophobic proteins present in the thylakoid membranes. Initial studies targeting the composition of the spinach and *Arabidopsis* thylakoid lumen were performed using 2-DGE analysis (18, 19). More than 60 soluble and peripheral membrane proteins were subsequently identified from purified pea thylakoid and detailed analysis of their targeting signals was performed (20). Thanks to the availability of the *Arabidopsis* genome annotation, a proteomic study targeting luminal and peripheral proteins (21) allowed identification of a total of 81 proteins using MS/MS and a new approach to predict the thylakoid lumen proteome *in silico*. This work was completed by several independent approaches (22, 23), leading to a consensus of more than 100 thylakoid luminal proteins (24, 25). Interestingly, these studies have shown that the chloroplast lumen proteins are not restricted to the generation of the pH gradient that fuels ATP synthesis, but also play important roles for the regulation of photosynthesis,

supporting protein turnover and protecting against oxidative stress.

The proteome of the thylakoid membranes has also been extensively studied. Initial MS-based studies of the thylakoid membrane proteins were essentially performed, in spinach and pea, on antennae or reaction-center subunits to identify the composition of the photosynthetic complexes and to detect post-translational modifications associated with these abundant proteins. Whitelegge and co-workers used ESI-MS to catalog intrinsic membrane proteins of the D1 and D2 reaction-center subunits from spinach thylakoids, to identify protein complexes components and to provide insights into native protein-protein interactions and their post-translational modifications (26). Furthermore, MS analysis of tryptic peptides released from the surface of *Arabidopsis* thylakoid membranes was used to characterize the reversible phosphorylation of chloroplast thylakoid proteins (27). These studies confirmed earliest data demonstrating that various subunits of the PSII and light-harvesting polypeptides (LHCII)¹ are phosphorylated. Some of these phosphorylation events were also found to be reversible in response to light or dark transitions. Zolla and co-workers also studied the light-harvesting proteins (LHCI or LHCII) from various monocot and dicot species and determined their intact molecular weights (28, 29). Whitelegge and co-workers also coupled ESI-MS with reverse-phase chromatography to catalog all detectable proteins in samples of PSII-enriched thylakoid membrane subdomains (grana) from pea and spinach (30). Only 30 proteins were identified, with important information on the phosphorylation of several PSII subunits. Later, the same group reported a set of 58 nuclear-encoded thylakoid membrane proteins from four plant species (31) and assigned experimentally the N-termini of all these proteins. This allowed testing the reliability of the different existing tools to predict chloroplast localization and/or cleavage sites starting from experimentally identified transit peptides. The first in-depth analysis of the thylakoid membrane was published by van Wijk and coworkers (32), resulting in the identification of 154 proteins. The same group later identified more than 240 thylakoid membrane proteins, of which, 86 were unknown (33). A recent study of thylakoid membrane dynamics, especially on environmentally modulated phosphoproteome induced by environmental changes was targeted to the photosynthetic membranes of *Chlamydomonas reinhardtii* (34). This study revealed

¹ The abbreviations used are: LHCI, light harvesting complex I; BBY, PSII enriched particles; CCD, Charge-Coupled Device; Chl, Chlorophyll; Col-0, Columbia 0; ECL, enhanced chemiluminescence; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; K, Kelvin; LAM, stroma-lamellae; LED, light emitting diode; LHCII, light harvesting complex II; LogFC, Log₂ of fold change; M, molar; MAR, margin; mM, millimolar; Nm, nanometers; PMSF, phenylmethylsulfonyl fluoride; PSI, photosystem I; PSII, photosystem II; SDS-PAGE, Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis; TBS, Tris-buffered saline; w/v, weight/volume; v/v, volume/volume.

that major changes in phosphorylation are clustered at the interface between the PSII core and its LHCII antennae. These data also suggest that the controlling mechanisms for photosynthetic state transitions and LHCII uncoupling from PSII under high light stress allow thermal energy dissipation. In a subsequent study also relying on intact mass measurements of membrane proteins, Zolla and co-workers analyzed all PSI and LHC proteins in ten different plant species, and identified PSI proteins present within stroma-lamellae of the thylakoid membrane (35). More recently, we performed large-scale analyses that aimed at identifying the whole chloroplast proteome. In this context, we focused, in the same set of experiments, on the localization of proteins in the stroma, thylakoids, and envelope membranes (14). The partitioning of each protein in these three chloroplast compartments was assessed using a semiquantitative proteomics approach (spectral counting). These data, together with an in depth investigation of the literature, were compiled to provide accurate subchloroplastic localization of previously known and newly identified chloroplast proteins. Among the 500 proteins that were identified in the thylakoid fraction, 220 could be assigned to the thylakoid compartment (see AT_Chloro database at http://www.grenoble.prabi.fr/at_chloro/).

Although proteomic analysis of thylakoids has allowed defining their protein composition, post-translational modifications and localization in the membrane and lumen compartments, a complete survey of the protein composition of the thylakoid subcompartments is still lacking. The objective of the present work was to complete previous studies and gain further information about the protein segregation in the grana and stroma-lamellae domains. To this aim, we have performed a complete survey of the protein composition of these thylakoid subcompartments using a semiquantitative proteomic approach to validate the differential distribution of thylakoid proteins between stroma-lamellae and grana. In this study, we not only confirm previous information about the localization of photosynthetic proteins, but also observe unexpected protein composition in photosynthetic complexes localized in both the BBY and the stroma-lamellae. This allows us to generate hypotheses about the assembly and function of thylakoid proteins.

MATERIALS AND METHODS

Plant Growth Conditions—*Arabidopsis thaliana* (ecotype Columbia, Col-0) were grown in culture chambers at 22 °C with a light intensity of 100 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (10-h light cycle) for 4–5 weeks before harvesting.

Purification of Intact Chloroplasts and Thylakoid Sub-compartments—All operations were carried out at 0–5 °C. Percoll-purified chloroplasts were obtained from 100–200 g of *A. thaliana* leaves as previously described (36). Purified intact chloroplast were lysed to retrieve thylakoids membrane in hypotonic medium in the presence of protease inhibitors (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.8, 0.2 mM benzamidine, 1 mM acid ϵ -amino caproic, 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and 10 mM

MgCl_2). Thylakoids membrane were centrifuged at $9700 \times g$ for 10 min and resuspended in the same buffer. Chlorophyll was extracted in 80% (v/v) acetone, and its concentration determined according to Porra *et al.* (37).

Margins and Stroma-lamellae Purification—Thylakoids at 0.5 mg chlorophyll/ml were incubated with 0.5% (w/v) digitonin (Sigma) for 30 min at 4 °C under agitation in order to solubilize the lighter fraction of the membranes. The mixture was diluted 10 times in the same hypotonic medium and subsequently centrifuged at $6000 \times g$ for 5 min. The pellet, consisting of nonsolubilized thylakoids, was collected and stored at 4 °C for further BBY purification while the supernatant was centrifuged at $15,000 \times g$. Again, the pellet was collected for the subsequent BBY purification whereas the supernatant was centrifuged again at $70,000 \times g$. The pellet, enriched in grana-margins was stored and resuspended in 10 mM HEPES, 5 mM MgCl_2 , 0.4 M sorbitol (in the presence of protease inhibitors), and stored in liquid nitrogen. The remaining supernatant was centrifuged at $140,000 \times g$ to pellet the stroma-lamellae fraction of the thylakoids, which was stored in the same buffer.

BBY Purification—Membranes, consisting of Pellet 1 (nonsolubilized thylakoids) plus Pellet 2 (grana obtained after solubilization with digitonin) were washed in resuspension buffer and diluted at 1 mg chlorophyll/ml concentration. The sample was incubated with Triton X-100 at 10 mg/mg chlorophyll for 30 min, in the dark, at 4 °C, under agitation. The mixture was centrifuged for 5 min at $3500 \times g$ at 4 °C. The supernatant was collected and centrifuged at $40,000 \times g$ for 30 min, at 4 °C. The pellet, consisting of BBY (*i.e.* the inner part of the grana stacks) was washed in resuspension buffer and centrifuged at $40,000 \times g$ for 30 min, at 4 °C. Membranes were stored in liquid nitrogen in 10 mM HEPES, 5 mM MgCl_2 , and 0.4 M sorbitol (in the presence of protease inhibitors).

Oxygen Evolution Measurements—Oxygen evolution was measured with a Clark-type electrode (Hansatech, UK). The reaction medium contained in a total volume of 1.5 ml 0.4 M sorbitol, 10 mM HEPES-KOH pH 7, 5 mM MgCl_2 , 0.2 mM PMSF, 0.2 mM benzamidine, and 1 mM ξ -amino caproic acid. Exogenous electron acceptors (3.5 mM ferricyanide and 250 μM 2,5-dichloro-p-benzoquinone) were added before measuring activity as evolution of molecular oxygen (38). Thylakoids or purified fractions (BBY and stroma-lamellae) were added at a final concentration of 12 $\mu\text{g chl}^{-1}$.

Solubilization Test for Thylakoid Membrane Proteins—To test selective solubilization of membrane proteins, freshly isolated thylakoid membranes were incubated for 30 min with the same detergents (digitonin or TRITON X-100) employed to isolate the stroma-lamellae and BBY fractions. Three different concentrations corresponding to (1) one half, (2) the same, and (3) the double concentration used in the purification protocol (see above) were tested in a volume of 100 μl containing 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.8, 0.2 mM benzamidine, 1 mM acid ϵ -amino caproic, 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and 10 mM MgCl_2 . At the end of the incubation (30 min at 4 °C), samples were centrifuged, and the pellet and supernatant fractions were collected. Proteins in the pellets were recovered in the same volume as the supernatants and stored in 10 mM HEPES, 5 mM MgCl_2 , and 0.4 M sorbitol (in the presence of protease inhibitors). The two fractions were then analyzed by SDS-PAGE and Western blots.

SDS-PAGE and Western Blot Analysis—Proteins were separated by SDS-PAGE as described by Chua (39). For Western-blot analyses, proteins were transferred to a nitrocellulose membrane (BA85, Schleicher & Schuell, and Whatman) which was blocked with 5% (w/v) dried-milk in Tris-buffered saline (TBS) containing 0.5% (v/v) Triton X-100. To assess the purity of the samples, antibodies raised against proteins associated to specific chloroplast subcompartments were used: AtpB (stroma-lamellae marker, Agrisera Vannas, Sweden)

at a 1:10,000 dilution, PsaD (stroma-lamellae marker, Agrisera) at a 1:3000 dilution, PsaC (stroma-lamellae marker, Agrisera) at a 1:10,000 dilution, PsaB (stroma-lamellae marker, Agrisera) at 1:5000 dilution, CP43 (BBY marker) at a 1:10,000 dilution (kindly provided by Dr. Luca dall'Osto, University of Verona), HMA1 (chloroplast envelope marker) at a 1:1,000 dilution (40), PHT4;4 (chloroplast envelope marker) at a 1:1000 dilution (41), and KARI (stroma marker) at a 1:5,000 dilution (kindly provided by Dr. Renaud Dumas, University of Grenoble). Proteins were detected by ECL.

77K Fluorescence Emission—To measure fluorescence emission spectra at cryogenic temperatures, a small (10 μ l) aliquot of the purified fraction was taken and loaded in the cuvette of a CCD spectrophotometer (JbeamBio, France) to measure fluorescence emission at 77 K. Fluorescence was excited with a blue LED peaking at 470 nm, and emission was recorded between 650 and 800 nm using a diode array. Spectra were normalized to the maximum peak (PSII for the BBY and PSI for the thylakoids and stroma-lamellae, respectively).

Protein Digestion—For MS analysis, 4 μ l of sample were loaded on a 12% SDS-PAGE Protein extracts were concentrated on a band, between the stacking and the separating gels (41). (7-cm gels, Bio-Rad, 4.5% stacking gel). Protein content varied between 2.6 and 9.3 μ g for the stroma-lamellae, 1.7 and 2.8 μ g for the thylakoids, and between 2.5 and 8.2 μ g for the BBY. The gel bands were manually excised and cut in pieces before being washed by six successive incubations of 15 min in 25 mM NH_4HCO_3 and in 25 mM NH_4HCO_3 containing 50% (v/v) acetonitrile using a Freedom EVO150 robotic platform (TECAN Lyon, France). Gel pieces were then dehydrated with 100% acetonitrile and incubated for 45 min at 53 °C with 10 mM DTT in 25 mM NH_4HCO_3 and for 35 min in the dark with 55 mM iodoacetamide in 25 mM NH_4HCO_3 . Alkylation was stopped by adding 10 mM DTT in 25 mM NH_4HCO_3 and mixing for 10 min. Gel pieces were then washed again by incubation in 25 mM NH_4HCO_3 before dehydration with 100% acetonitrile. 0.15 μ g of modified trypsin (Promega, Madison, WI; sequencing grade) in 25 mM NH_4HCO_3 was added to the dehydrated gel pieces for an overnight incubation at 37 °C. Peptides were then extracted from gel pieces in three 15 min sequential extraction steps in 30 μ l of 50% (v/v) acetonitrile, 30 μ l of 5% (v/v) formic acid, and finally 30 μ l of 100% acetonitrile. The pooled supernatants were then dried under vacuum.

Nano-LC-MS/MS Analyses—All the thylakoid fractions were run as biological triplicates and technical duplicates. The dried extracted peptides were resuspended in 5% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid and analyzed by online nanoLC-MS/MS (Ultimate 3000, Dionex and LTQ-Orbitrap Velos Pro, Thermo Fischer Scientific Waltham, MA). Peptides were sampled on a 300 μ m x 5 mm PepMap C18 precolumn and separated on a 75 μ m x 250 mm C18 column (PepMap, Dionex). The nanoLC method consisted in a 120-min gradient at a flow rate of 300 nL/min, ranging from 5% (v/v) to 37% (v/v) acetonitrile in 0.1% (v/v) formic acid during 114 min before reaching 72% (v/v) for the last 6 min. MS and MS/MS data were acquired using Xcalibur V2.1.0 SP1 (Thermo Fischer Scientific). Spray voltage and heated capillary were respectively set at 1.4 kV and 200 °C. Survey full-scan MS spectra (m/z = 400–1600) were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 106 ions (maximum filling time: 500 ms). The twenty most intense ions from the preview survey scan delivered by the Orbitrap were fragmented by collision-induced dissociation (collision energy 35%) in the LTQ after accumulation of 104 ions (maximum filling time: 100 ms).

Database Searching—Peak lists were generated with the Mascot Distiller version 2.4.3.3 software (Matrix Science, Boston, MA) from the LC-MS/MS raw data. Using the Mascot 2.4.0 search engine (Mascot Daemon, Matrix Science), MS/MS spectra were searched against a compilation of the *A. thaliana* protein database provided by

TAIR (nuclear, mitochondrial, and chloroplast genomes; TAIR version 10; 35,386 entries; <http://www.arabidopsis.org/>) and a homemade list of contaminants frequently observed in proteomics analyses (260 sequences, (14)). Both databases were concatenated with the corresponding decoy database in order to calculate false discovery rate. One missed trypsin cleavage was allowed, and trypsin/P was used as the enzyme (cleavage at the C terminus of Lys/Arg). The mass tolerances were 10 ppm for precursor ions and 1.0 Da for fragment ions. The variable modifications allowed were N-acetylation (protein N termini), methionine oxidation and dioxidation. A fixed modification was set: carbamidomethyl cysteine. Mascot search results were automatically filtered using the home-developed IRMa 1.31.1 software (42). The following parameters were applied: (1) the number of report hits was fixed automatically to retrieve proteins with a p value, as defined by Mascot, such as $p < 0.05$; (2) only peptides ranked first and with an homology threshold such as the p value, as defined by Mascot, corresponding to $p < 0.05$, were kept. Every duplicated peptide sequence was conserved to calculate spectral counts. Consistent protein grouping for all the analyses (same master protein reported) was ensured using the IRMa software. In order to avoid redundancy, only the master proteins (i.e. the protein representing a protein group) were given in the provided results. For each database search, protein groups were identified with a 1% false discovery rate (14). Database searching results and raw files were exported in the PRIDE format and deposited to the ProteomeXchange server with identifier PXD000546 and DOI 10.6019/PXD000546 and PRIDE accession # 31943 (www.proteomexchange.org). Lists of proteins and peptides can also be found in the Supplemental Data file. Database searching results, with annotated spectra (Pride_Exp_Complete_Ac_31943.xml), can be visualized using the freely available PRIDEInspector tool (<https://code.google.com/p/pride-toolsuite/wiki/PRIDEInspector>).

Data Processing—Total spectral counts and spectral counts for proteotypic peptides were retrieved from the IRMa filtered results for each of the MS/MS analyses. Weighted spectral counts were calculated as described in Abacus (43) and were imported in R (version 3.0.1) (44), which is an open source statistical analysis environment. The raw quantitative data were converted into a “MsnSet” class object from the R package MSnbase (44 and 45). The MsnSet class is derived from the “eSet” class and mimics the “ExpressionSet” class used for microarray data. An “MsnSet” object contains quantified expression data for MS proteomic data as well as the experimental meta-data, and allows the analysis of quantitative MS-based data using advanced statistical tools implemented in R, and particularly that of the pRoloc R package (46) (<http://www.bioconductor.org/packages/2.12/bioc/html/pRoloc.html>), which takes as input objects of the MSbase classes. Raw quantitative data were subjected to some pre-processing steps (Fig. 1), including missing value imputation, log-transformation and sample normalization in order to render it compliant with the basic standards of good quality data (completeness, consistency and accuracy). Note that within the whole pipe-line dedicated to the quantitative data analysis, all type of replicates (biological or technical) for a given group (e.g. the BBY group) have been treated equally.

Missing Value Imputation—Quantitative proteomics data are well known for high rates of missing values. This fact requires deeper consideration when trying imputing them. In the present study, 1295 proteins were identified in four groups of samples (thylakoids, stroma-lamellae, margins, and BBY). However, not all proteins were identified in each individual group. For such proteins, the quantitation algorithm associated a NA (nonattributed) value to all samples where that protein was not identified. The actual missing values (MVs) were recorded for proteins that had been identified in at least one sample in a particular group, and they were also assigned with NAs. The proteins concerned are unlikely to be absent in a particular biological compartment as they have already been identified in at least one

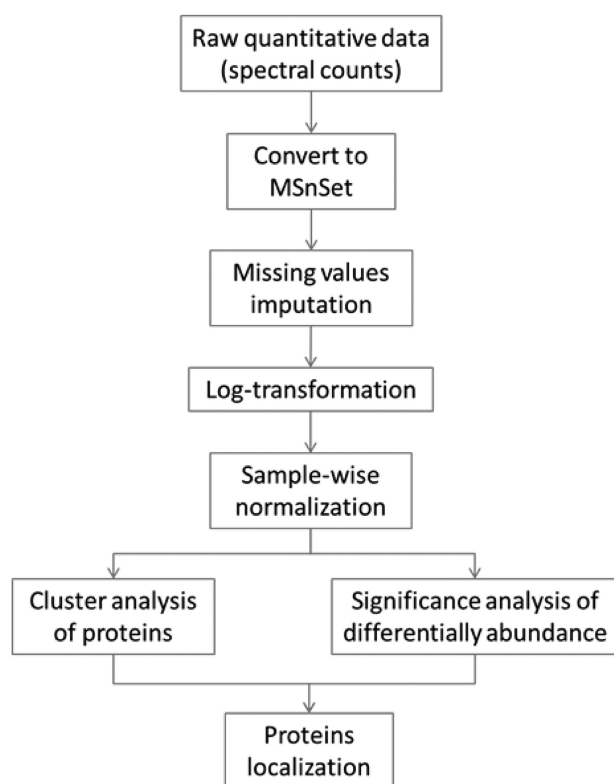


FIG. 1. Pipeline for data analysis.

sample. According to (47), in quantitative proteomics, MVs can be explained by several potential mechanisms: (1) the protein/peptide is present at abundance the instrument should detect it, but it is not detected or incorrectly identified; (2) the protein/peptide is present at an abundance that the instrument cannot detect them; and (3) the protein/peptide is not present at all. MVs recorded as a result of (2) and (3) are called censored missing values, and it has been shown that their simple imputation based on observed values is not appropriate, leading to an overestimate of peptides abundance and to biased results (48). Conversely, if the MVs are recorded according to 1) (the so-called Missing Completely at Random - MCAR), one can use the observed values for their imputation. Thus, depending on the nature of the MVs, different strategies should be employed for their imputations. For MCAR data one can use simple MVs imputation methods which rely on the observed values (48), whereas for the censored data one strategy is replace the MV with the minimum value observed (47). Our strategy for MV imputation was based on the above considerations. Although different probabilistic models are proposed to separate between MCAR data and censored data (49, 50), our approach is based on a statistical investigation of the data in order to verify whether the missing values are abundance dependent or not. To do so, in each group of samples and for each protein we estimated the mean value of the nonmissing values, and we visualized the distribution for all proteins. This provided useful information that helped infer whether the missing values are abundance dependent or not and provided guidance in declaring the missing values as being censored or MCAR. The MVs imputation strategy consisted of replacing the censored MVs by the minimum value observed, while the MCAR MVs they were imputed using a nearest neighbor strategy (51), which was implemented as a routine preprocessing in MSnset (45) (Supplemental Material).

Data Normalization—For the 24 samples corresponding to the 4 biological groups (thylakoids, BBY, margins and stroma-lamellae), the

distributions of the log-transformed abundances were examined in order to investigate about the presence of technical bias. For each of the 4 groups of samples, quantiles normalization (52) was performed after the imputation of missing values implemented in the MSnSet R package (Supplemental Material).

Statistical Analysis of Differential Abundance—We employed the R package LIMMA (53) which implements a robust method for differentially abundance assignment widely used for microarray data, based on linear protein-wise models as well as an empirical Bayesian model for robust estimates of protein-wise standard errors. As the MSnSet class mimics the “ExpressionSet” class used for microarray data, and as this latter is used as input object for LIMMA, it appears that any proteomics dataset converted into the MSnSet format is compliant with a LIMMA use. The method in LIMMA is based on a modified version of Student *t* test (called moderated *t* test), where the standard errors are reduced toward a common value using a simple Bayesian model. This means that, when estimating the variable-wise standard error, the model borrows the information from the other genes (or in this case, proteins), which leads to more robust estimates. As a consequence, the model is well-suited to assess differentially abundance for datasets with few samples. In a recent comparison study on the significance testing for small microarray experiments (54), LIMMA has been found as being the method that balances the best the number of true positives and false positives. Within this comparison study, the authors used experiments with small sample size ranging from 2 to 5 samples per group. The approach consists in first fitting a linear model to each protein and second, estimating the corresponding moderated *t*-statistic using the empirical Bayes procedure. *p* values were calculated using the moderated *t*-statistic. Multiple testing corrections were then performed by adjusting the *p* values for multiple testing using Benjamini and Hochberg’s method to control the false discovery rate (55). We set the adjusted *p* value (referred to as *q*-value) threshold for declaring a protein differentially expressed to 0.05, meaning that the expected proportion of false discoveries in the selected group of proteins is controlled to be less than 5%.

Proteins Cluster Analysis—For each group of samples (BBY, margins and stroma-lamellae) we derived a representative sample as being the mean over the six available replicates divided by the mean abundance values across the six thylakoid samples, in order to take into account for the difference in abundance of thylakoid proteins. Working with such relative abundance values will homogenize low and high abundant proteins, providing that proteins more abundant in BBY with respect to stroma-lamellae will not be sensitive to the actual protein abundance. As a consequence, a number of 255 proteins, which were not identified in the thylakoids samples, were not considered. Cluster analysis was performed using PAM (Partition Around Medoids) (56), a particular implementation of the *k*-medoids algorithm. *k*-medoid is one of the numerous variations of the extremely well-known *k*-means algorithm. This latter is implemented in the R package pRoloc (46) (<http://www.bioconductor.org/packages/2.12/bioc/html/pRoloc.html>), as well as numerous more sophisticated algorithms, such as kernel *k*-means, spectral clustering (57) (by mapping to the corresponding R package (58)), or hyperspherical clustering (59) (by direct implementation). As methods to automatically assess the right number of clusters are still under development for the next version of pRoloc, this point was tackled as described in Supplemental_Material.

Curated Annotation of Proteins—Information about identified proteins were retrieved and compiled from the literature (<http://www.ncbi.nlm.nih.gov/pubmed>) using AGI numbers or protein names and screening for recently published data, from UniProtKB (<http://www.uniprot.org/>), TAIR (<http://www.arabidopsis.org/>), PPDB (60) (<http://ppdb.tc.cornell.edu/>), AT_Chloro (14) (http://www.grenoble.prabi.fr/at_chloro/), SUBA3 (61) (<http://suba.plantenergy.uwa.edu.au/>) and

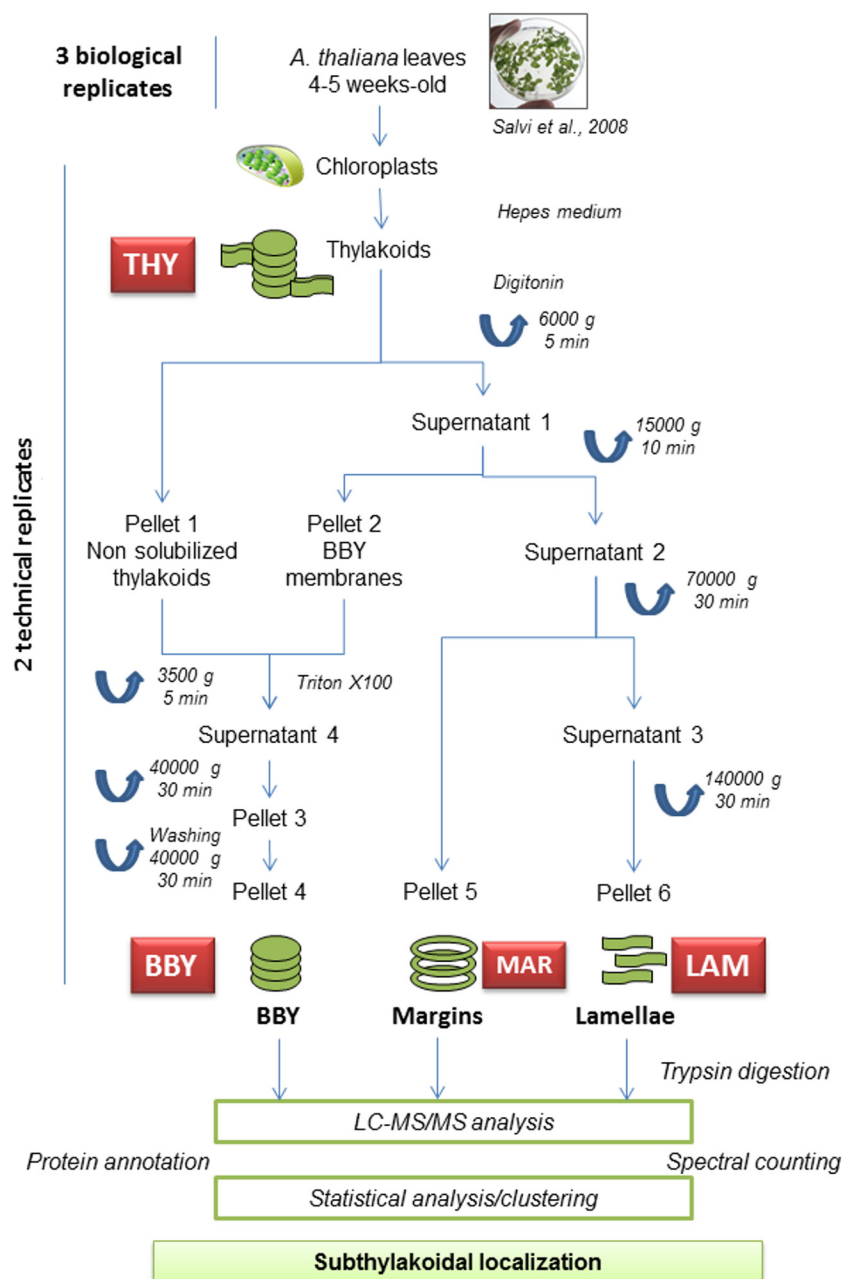


FIG. 2. Purification of chloroplast subfractions: overview of the complete strategy. Chloroplasts from *Arabidopsis* leaves were purified on Percoll density gradients. The stroma-lamellae, margins and BBY fractions were purified using different detergents and sequential centrifugations.

the ontological tool MapMan (62) (<http://mapman.gabipd.org>) which was specifically designed to cover plant-specific pathways and processes (63).

RESULTS

Optimized Thylakoid Fractionation to Generate Sub-thylakoidal Proteomes: Overview of the Strategy—Purification of thylakoid subfractions (Fig. 2) started from Percoll-purified intact chloroplasts from *A. thaliana* leaves (36). The rationale for this approach was to minimize contamination by mitochondria, which are largely removed from the sample during the isolation

of intact chloroplasts. After chloroplast lysis in a hypotonic medium, thylakoids were solubilized with digitonin (10 mg/mg Chl). After centrifugation at $15,000 \times g$, the pellet was used for the purification of the inner grana fraction (BBY), following the protocol established by Berthold *et al.* (38). The supernatant was washed, and centrifuged at $140,000 \times g$ to recover stroma-lamellae (Fig. 2). Starting from 100–200 g of *Arabidopsis* leaves, we obtained 40 μ g of chlorophyll for the stroma-lamellae and 80 μ g of chlorophyll for the BBY (corresponding to $\sim 250 \mu$ g and $\sim 200 \mu$ g of proteins respectively).

To obtain statistically reliable results, three biological replicates were prepared and two technical replicates were generated for each chloroplast biological replicate. Therefore, six samples were produced for each of BBY, stroma-lamellae (LAM), margin (MAR) and thylakoid (THY) fractions during the purification process (Fig. 2). All the protein samples were concentrated on a SDS-PAGE between the stacking and the separating gels according to (41) prior to LC-MS/MS analyses that were performed using a LTQ-Orbitrap VelosPro coupled to a nano-LC system (Fig. 2). As previously described in the context of the subchloroplastic localization of proteins (14), spectral counting was used for relative quantification in order to determine the subthylakoid localization of identified proteins. Different spectral counting measures can be obtained for a given identified protein group: total number of MS/MS spectra corresponding to the identified peptides, the number of MS/MS spectra corresponding to proteotypic peptides, or weighted spectral counts. In order to accurately deal with peptides shared between different protein groups, weighted spectral counts were used. Indeed, it was shown that distributing spectral counts of shared peptides, based on the presence of unique peptides in a given protein group, generated the best results (64). Consequently we chose to calculate weighted spectral counts as suggested in Abacus (43) and used these metrics for subsequent statistical analysis. Then, both statistical and clustering analyses were performed in order to identify proteins that were significantly differentially distributed between the different subthylakoid fractions. In addition, all identified proteins were thoroughly annotated with respect to their known or expected subcellular and subchloroplastic localizations and function for further data mining purposes. Eventually, 1295 proteins were identified from all the analyses performed (supplemental Table S1).

Evaluation of Cross-contaminations at the Sub-cellular, Sub-chloroplastic, and Sub-Thylakoid Levels—Cross-contaminations at the subcellular level was evaluated using analysis of the 1295 proteins identified from all the analyses performed and manual annotation or prediction of their subcellular localization (supplemental Table S1). In good agreement with proteomic analyses previously performed on compartments derived from Percoll-purified chloroplasts (14), spectral counting analysis of our data indicate that cross-contamination with other subcellular compartments is very low (*i.e.* less than 5%; supplemental Fig. S1).

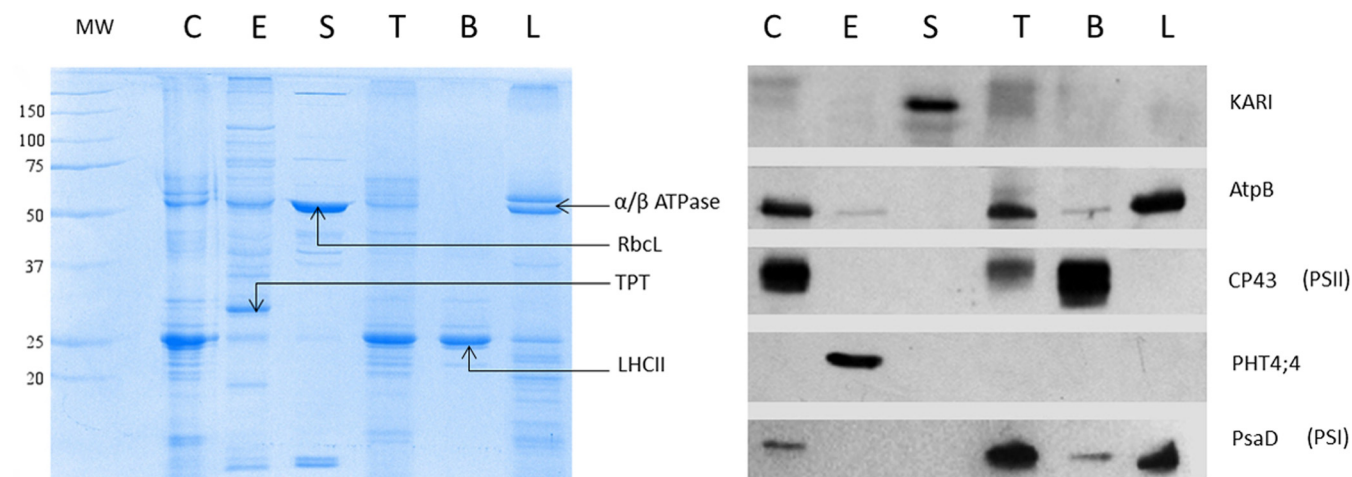
We also quantified the cross-contamination of the subthylakoid fractions (stroma-lamellae and BBY) using marker proteins associated to the stroma and envelope subcompartments, the two other main subchloroplastic compartments. This was done using specific antibodies raised against marker proteins from the different subchloroplastic fractions (the large subunit of RuBisCO, a known marker from the stroma and PHT4;4, a phosphate transporter associated to the inner membrane of the chloroplast envelope). For each analysis, three different biological replicates were tested. The level of

cross-contamination was qualitatively evaluated by direct comparison of the intensities of the signals (Western-blot) arising from the fractions to be tested and those arising from a range of dilutions between 3 and 100% of purified fractions derived from other chloroplast compartments. Data reported in supplemental Fig. S2 indicate an average level of cross-contamination between 3 and 20% for all the purified fractions analyzed. Overall, stroma and envelope contaminations appeared to be lower than 3% in the BBY fraction, whereas a larger contamination by envelope proteins (between 10–15%) was detected in the stroma-lamellae fraction. Contamination of the stroma-lamellae fraction with the marker from the stroma was found to be lower than 10%.

We next assessed the specific protein composition of subchloroplastic fractions (Fig. 3) using SDS-PAGE and Western-blot analyses targeting known markers of the various purified subfractions. As shown in Fig. 3A, a Coomassie blue staining of the proteins present in the different fractions and separated on a SDS-PAGE identifies different polypeptide profiles, when compared with that of intact chloroplasts or isolated thylakoids. Indeed, as expected, LHCII (Fig. 3A, LHCII), the major antenna complex of the PSII, was observed to be highly enriched in the BBY fraction (Fig. 3A, lane B) while it was barely detectable in the stroma-lamellae fraction (Fig. 3A, lane L). Subunits of the ATP synthase complex (Fig. 3A, α/β ATPase), which are known to be present in the nonappressed regions of the thylakoids, could not be detected in corresponding lane of the SDS-PAGE (Fig. 3A, lane B) consistent with the absence of this complex in the grana stacks. Both in BBY and stroma-lamellae fractions, we barely detected known abundant markers associated to other chloroplast subcompartments, such as the RuBisCO (Fig. 3A, RBCL) for the stroma (Fig. 3A, lane S) or the phosphate/triose-phosphate transporter (TPT) for the envelope (Fig. 3A, lane E).

The enrichment of specific complexes in purified stroma-lamellae and BBY fractions was further confirmed by Western-blot analyses, using specific antibodies. Fig. 3B shows that CP43, a chlorophyll protein associated with PSII, was largely enriched in the BBY fraction (Fig. 3B, lane B) when compared with the thylakoids (Fig. 3B, lane T) or the intact chloroplasts (Fig. 3B, lane C) fractions. By contrast, no CP43 could be detected in the stroma-lamellae fraction (Fig. 3B, lane S). AtpB and PsdA, components of the ATP synthase CF₀-F₁ and of PSI respectively (two complexes known to be predominant in the stroma-lamellae), were found both in the thylakoids (Fig. 3B, lane T) and in the chloroplasts (Fig. 3B, lane C). AtpB and PsdA were enriched in the stroma-lamellae (Fig. 3B, lane S), while being largely diminished in the BBY fraction (Fig. 3B, lane B).

As our objective was to obtain accurate data about subthylakoid localization, we then assessed the presence of stroma-lamellae-enriched complexes in the BBY, and *vice versa*, using the same Western-blotting approach (supplemental Fig. S2). We found that > 10% of the PsdA subunit was localized



A FIG. 3. Characterization of Percoll-purified *Arabidopsis* chloroplast and subplastidial fractions. C, crude chloroplast proteins; E, envelope membrane proteins; S, stroma proteins; T, thylakoid proteins; B, BBY proteins; and L, stroma-lamellae proteins. Each fraction contained 6 μ g proteins. A, Fractions were analyzed on a 12% SDS-PAGE. α/β ATPase: alpha and beta subunit of the ATPase (stroma-lamellae marker), RbcL: Large subunit of RuBisCO (stroma marker), TPT: Phosphate-triose-phosphate translocator (envelope marker), and LHCII: Light harvesting complex proteins (BBY marker). B, Western-blots were performed with using polyclonal antibodies raised against KARI (At3g58610, stroma marker), AtpB (AtCg00480) and PsaD (At4g02770) (stroma-lamellae markers), CP43 (AtCg00280, BBY marker), and Pht4;4 (At4g00370, envelope marker).

in the BBY. However, numbers became smaller in the case of another PSI subunit (supplemental Fig. S2; lanes PsaD and PsaC), suggesting a possible heterogeneity at the level of this complex (see below). On the other hand, less than 3% of the PSII complexes (supplemental Fig. S2, lane CP43) were located in the stroma-lamellae. This indicates that, as expected from previous studies (for a review, see (65)), none of the major photosynthetic complexes were exclusively localized in a single compartment (with the possible exception of the ATPase). However, a clear enrichment of the major complexes in one compartment is seen. This was further confirmed by measuring the fluorescence spectra of the different fractions at cryogenic temperatures (Fig. 4). At variance with room temperature, PSI fluorescence becomes visible at 77K. Therefore, three main fluorescence emission peaks were detected in the red region of the visible spectrum, upon excitation of purified thylakoids with blue light. The first two peaks (~ 685 and ~ 695 nm) represent fluorescence emission from PSII, while the large fluorescence peak seen at ~ 730 nm stems from PSI. The two PSII peaks were barely detectable in the purified stroma-lamellae fraction and the PSI fluorescence peak alone turned out to be extremely reduced in the purified BBY fractions when compared with the thylakoids (Fig. 4). In no case, was emission from energetically uncoupled chlorophyll observed, indicating that no damage to the two photosystems was induced by our purification protocol. Measurements of the chlorophyll a/b ratio and oxygen evolution capacity of the different fractions (supplemental Table S2) also confirmed the successful separation of the two fractions. The chlorophyll a/b ratio of BBY fraction turned out to

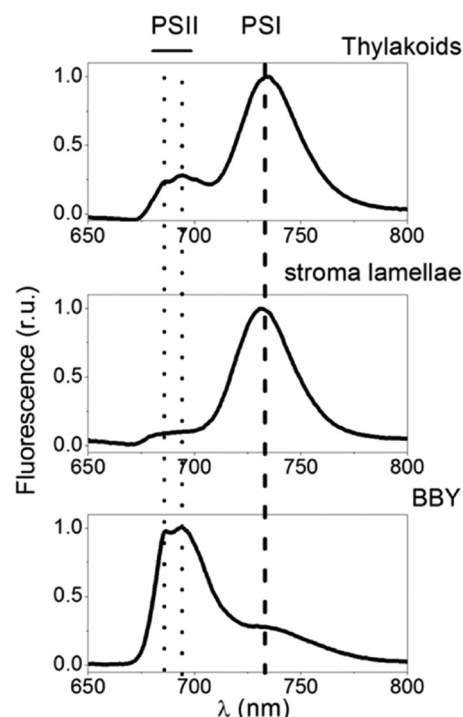


FIG. 4. 77K chlorophyll fluorescence emission spectra of thylakoid membrane subfractions isolated from *A. thaliana*. Thylakoid proteins (or thylakoid subfractions) samples were loaded on to a metal cuvette, which was directly bathed into a liquid nitrogen solution. Fluorescence spectra were recorded upon excitation at 470 nm. Peaks were normalized for comparison purposes. Top spectrum: isolated thylakoids. Medium spectrum: stroma lamellae. Bottom spectrum: BBY fraction.

be much lower (1.7) than that of thylakoids and stroma-lamellae (2.9 and 4.2 respectively) consistent with the well-established notion that chlorophyll *b* is accumulated in the PSII antenna (66) because it is mainly contained in the LHCII complexes. Measurements of oxygen evolution in presence of exogenous electron acceptors (ferricyanide and 2,5-dichloro-*p*-benzoquinone) showed that our BBY preparation retained most of the O₂ evolution capacity of the starting material (thylakoids) while no oxygen evolution was detectable in the stroma-lamellae fraction. This confirmed the large enrichment of PSII in the BBY preparations when compared with the stroma-lamellae.

In conclusion, Western-blot, cryogenic temperature fluorescence, oxygen evolution and assessment of the chlorophyll *a/b* ratios analyses indicate that cross-contaminations between stroma-lamellae and BBY fractions, or their contamination by other chloroplast compartments were low enough to justify a MS-based analysis of the subthylakoid localization of the proteins embedded in the photosynthetic membranes.

An Updated Repertoire of Chloroplast and Thylakoid Proteins—The present study allowed the identification of 1295 proteins (supplemental Table S1). We first evaluated the coverage of the chloroplast thylakoid proteome by combining present data with earlier large-scale analyses targeted to the chloroplast. To this aim, we compared proteins identified during this work with those previously reported by Ferro *et al.* (14) (AT_CHLORO database) and Zybailov *et al.* (15), that is, two extensive studies, performed at the whole chloroplast level, where similar numbers of proteins were identified (supplemental Fig. 3). These previous chloroplast data and present results were collated to obtain a total of 2103 nonredundant proteins. The overlap of the present study with the other data shows that the AT_CHLORO database contained more than 300 proteins that were not identified during the study of Zybailov *et al.* (15). This could be easily explained by the specific enrichment (a factor of ~50) resulting from the purification of the envelope fraction that was specific to the study of Ferro *et al.* (14) and allowed detection of minor chloroplast envelope proteins.

When compared with previous large scales proteomic analyses targeting the chloroplast, of the 1295 proteins identified in the thylakoid fractions, 361 (28%) novel proteins were identified in the present study (Supplemental Table S3). Most of them were identified with some of the lowest spectral count measures (supplemental Table S3, columns R, S, T, and U), and often detected in only one of the fraction (supplemental Table S3, columns D, E, F, and G). In order to estimate the proportion of the proteins that are genuine chloroplast proteins we analyzed their subcellular and subchloroplastic localization using data from the literature and targeting prediction tools (supplemental Table S3, columns L, M, N, and O). About 70 proteins, which are known to be major components of other cell compartments (plasma membrane, cytosol, nucleus, Golgi, vacuole, mitochondria, peroxisome), were clas-

sified as cross-contaminants (supplemental Table S3, columns H and I, “Others”). However, more than one third of these 361 proteins are either known or predicted (using ChloroP and or TargetP tools) to be targeted to the chloroplast, and few more proteins, while lacking a predicted targeting peptide, are genuine chloroplast proteins (supplemental Table S3, columns H and I, “C ?”). Finally, many other proteins (~ 150), while also lacking a predicted targeting peptide, were never detected in another cell compartment using previous proteomic approaches targeting the other cell compartments cited above. Thus, some of these proteins (supplemental Table S3, columns H and I, “C ??”) may clearly be genuine chloroplast components that could be detected here thanks to the combination of 1) thylakoid membrane fractionation (enrichment of minor components) and 2) the improved sensitivity of state-of-the-art mass spectrometers. Among those 361 proteins only 14 (At3g27690, At3g21055, At1g51400, At2g26500, At1g18730, At4g14870, At3g56010, At5g02160, At4g38100, At3g17930, At3g26580, At2g45180, At5g46390, and At3g63540) were previously identified in studies aimed at the characterization of thylakoid proteomes (supplemental Table S4).

We next compiled the main thylakoid proteome studies which correspond to four types of investigations targeting different thylakoid fractions: thylakoid membranes (32, 33), thylakoid lumen (21, 22), whole thylakoid (14) and BBY/stroma-lamellae (present work). A set of 1400 proteins was collated (supplemental Table S4). In order to determine what is the minimal thylakoid proteome, as defined from proteomics experiments, successive different levels of curation were applied. Based on present curated annotations, we removed contaminant proteins (supplemental Table S4, protein labeled as “na” in the “Main subplastidial localization” column). Referring to MapMan classes (62), we also removed proteins that belong to classes for which proteins are acknowledged to be localized either in other chloroplast compartments, that is, stroma and envelope, or in other subcellular compartments. These MapMan classes are the following ones: Calvin Cycle, DNA, envelope transporters, glycolysis, mitochondria electron transport, mitochondria transporters, protein aa (amino acid) activation, protein synthesis (ribosomal proteins), RNA processing, RNA transcription, RNA.RNA binding, starch metabolism and TCA (tricarboxylic acid cycle proteins). Proteins that belong to the class “PS/light reaction” were labeled as thylakoid proteins. We also labeled as thylakoid proteins those that were identified in the thylakoid with a percentage ≥ 50% of occurrence in our AT_CHLORO database (14). In addition, proteins that were annotated in supplemental Table S1 to be Ch/Th were labeled as being thylakoid proteins. Proteins being annotated to be located in either in the stroma or the envelope were referred as nonthylakoid proteins. Remaining proteins were labeled “unknown.” Thus, from the present annotation of thylakoid proteomes, one can consider that ~ 300 proteins are true thylakoid proteins. Besides, 400 proteins are potentially thylakoid proteins but for those proteins

less biological evidences are available (supplemental Table S4). Therefore about half of the 1400 proteins collected can be considered to be contaminants either from other subcellular compartments or from other subchloroplastic fractions. Although this number seems important, it has to be relativized with respect to the actual amount of each protein identified in a given fraction. Moreover, when considering those proteomic analyses of thylakoid fractions, it is obvious that some contaminants from the stroma are difficult to avoid (e.g. highly abundant Calvin cycle and ribosomal proteins). Analysis of the compilation of major thylakoid proteomes indicate that 812 proteins were not identified in previous thylakoid proteomics investigations (supplemental Fig. S4). Among those 812 identified proteins, 25 were known to be located or partly located in thylakoids (e.g. several NADH DH proteins) and 281 proteins were annotated for not having determined subchloroplastic localization. The latter class of proteins is particularly interesting as it potentially gathers new thylakoid proteins. Indeed, cross-contamination analysis suggests that these proteins are very likely to be new chloroplast proteins as contamination from other subcellular compartments is very low (less than 5%) and results from abundant and well-known proteins (e.g. cytosolic ribosomal proteins, mitochondrial ATP synthase, vacuolar ATPases). However, the actual localization of those chloroplast proteins at the subchloroplastic level needs additional information to be collected.

Thus, from the comparison with chloroplast and thylakoid proteome identified during previous studies we decided to extract potential new thylakoid proteins. From the proteins specifically identified in the present study, compared with chloroplast and thylakoid proteomes (see above), we filtered out the most likely contaminants from other cell and chloroplast compartments. An eventual set of 218 proteins was retrieved (supplemental Fig. S5). We chose to perform a detailed analysis of those 218 potentially new thylakoid proteins that were identified during this work. Most of these proteins were never detected using recent large-scale proteomic analyses and are thus expected to be low abundant proteins in the chloroplast or in other cell compartments contaminating the purified thylakoid fractions. To classify these proteins, we first used targeting prediction tools (ChloroP and TargetP) and database annotations (SUBA, PPDB) for the analysis of their subcellular and subplastidial localization. We also screened the literature and other databases (TAIR, MapManBin, UniProtKB), to extract information about the available protein descriptions and functions (supplemental Table S5, rank 1–63). Out of these 218 proteins, 63 were predicted to be targeted to plastids using both ChloroP and TargetP targeting prediction tools, most of them (50 proteins) being thus also annotated as putative plastid proteins in the SUBA database. We thus considered that these 63 proteins were indeed good candidates for new chloroplast and thylakoid components.

We then carefully analyzed the remaining 155 proteins and selected an additional set of 40 proteins that could also be

associated to the list of these new chloroplast components, these proteins being predicted to contain a chloroplast transit peptide using either ChloroP or TargetP, or already classified as plastid protein in SUBA or, for very few proteins, experimentally associated to the chloroplast in more targeted studies (supplemental Table S5, rank 64–103). A detailed analysis (description, function, protein sequence, literature mining) of these proteins allowed supporting this classification. For instance, the YCF3 (AtCg00360) is chloroplast encoded and thus does not contain a predictable chloroplast transit peptide. The dynamin-like protein ARC5 (At3g19720), an outer chloroplast envelope protein, is, according to the predictions, also devoid of a chloroplast transit peptide) like most outer envelope proteins (67). *Arabidopsis* ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. As another example, a CAAX-like protease (At1g14270) did not appear in our list of ChloroP positive proteins. However, checking the other putative gene models in TAIR revealed that one of the CAAX-like gene models (At1g14270.1) is predicted to contain a chloroplast transit peptide using this ChloroP software.

Consequently, out of the 218 new proteins identified here, at least 103 of them (63 + 40) might be genuine thylakoid (or at least chloroplast) proteins. Very few of the remaining 115 proteins could be further considered for a tentative chloroplast localization because most of them are either totally unknown proteins or only contain domains that allow to suspect or predict a putative function such as proteases, phosphatases, kinases, transporters, TPR or PPR proteins (supplemental Table S5, rank 104–218).

In conclusion, bioanalysis of the present results together with previous chloroplast and thylakoid proteomic analyses indicate that a core of ~300 well-characterized thylakoid proteins have been identified in chloroplast or thylakoid proteomic studies so far. In addition, from this data collection we showed that we identified at least a hundred potential new thylakoid proteins as discussed below.

Statistical Analysis of Differential Abundances for Protein Localization and Proteins Clustering—The present work is the first study aimed to specifically address the accurate proteomic-based localization of thylakoid proteins, in the two major thylakoid substructures: BBY and stroma-lamellae (LAM). The main goal of the present study was to provide reliable data for a better understanding of the respective role of these compartments, especially with respect to PS protein complexes.

The discovery of proteins that are more enriched in BBY compared with stroma-lamellae (or *vice-versa*) raises the question of how it is possible to differentiate genuine grana and stroma-lamellae proteins from possible contaminants. The issue of protein differential abundance is similar to the discovery of differentially expressed genes, as revealed by microarray gene expression analyses (69). From the plethora

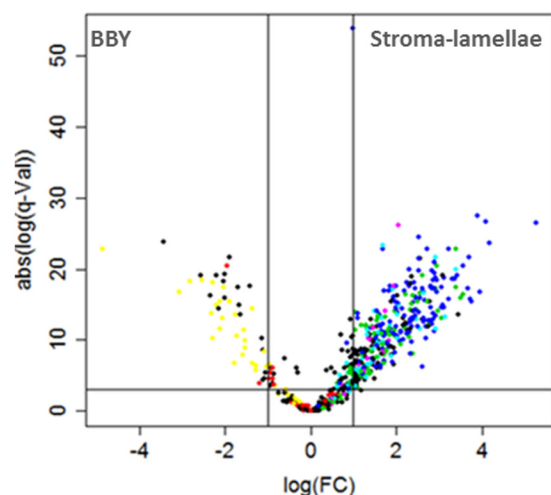


FIG. 5. **Volcano plot of quantified proteins for the BBY versus stroma-lamellae samples.** One dot corresponds to one protein; the color code corresponds to the labels from the proteins cluster analysis (see supplemental Fig. S5).

of hypothesis testing algorithms proposed, we decided to perform the differentially abundance analysis using LIMMA (53) (with the moderated t test and with an adjusted p value threshold (55) of $q = 0.05$ (see the Materials and Methods section for details). This results in a number of 515 proteins that are differentially distributed between grana and stroma-lamellae. Among them, the proteins with $\text{LogFC} > 0$ are more abundant in stroma-lamellae, whereas those with $\text{LogFC} < 0$ are more abundant in BBY (Fig. 5). Based on these results we could identify 65 proteins more abundant in BBY compared with stroma-lamellae ($q < 0.05$ and $\text{LogFC} < 0$) and 450 proteins more abundant in stroma-lamellae than in BBY ($q < 0.05$ and $\text{LogFC} > 0$) (supplemental Table S6).

The same approach was used to test an additional fraction that was purified in our samples, that is, the BBY margins and end membranes (MAR fraction; Fig. 2). This fraction is defined as the portion of the appressed thylakoids in contact with the stroma and, according to previous studies, represents an essential compartment for photosynthetic activity (70). When performing differentially abundance analysis between margin fraction samples and BBY samples we identified a number of 296 differentially abundant proteins ($q < 0.05$), among which 238 are more abundant in margins ($\text{LogFC} > 0$). However, most of them, 229 proteins ($\sim 96\%$), were also enriched in stroma-lamellae when compared with BBY. The remaining 59 differentially abundant proteins were more enriched in BBY samples ($\text{LogFC} < 0$), and among those 59 proteins, 51 ($\sim 87\%$) of them are also found as being more abundant in BBY when compared with stroma-lamellae samples. Overall, only 20 proteins are found as being more abundant in margins compared with stroma-lamellae samples ($\text{LogFC} < 0$), from which only five (At3g23700, AtCg00580, At5g01920, At5g67050, and At1g77090) appear not to be significantly distributed in BBY or stroma-lamellae when BBY and stro-

ma-lamellae were compared (supplemental Table S7). From these results we can conclude that the vast majority of proteins have similar representation in the stroma-lamellae and in the margins samples. Indeed the lists of differentially abundant proteins found as being more abundant in stroma-lamellae and margins respectively, when compared with BBY-enriched proteins, contain $\sim 96\%$ common proteins. In addition there are few proteins that are more abundant in margins than in stroma-lamellae. Two explanations can be proposed for this finding. First, the purification of “pure” margin fraction is an extremely difficult task, because of their physico-chemical properties, which are very similar to those of the stroma-lamellae (65). Second, margins are likely to have a protein composition that strongly reminds that of the stroma-lamellae. Being nonappressed regions of the grana, the margins are likely to host proteins that cannot be retrieved in the stacks because of their bulky stromal exposed parts. Therefore, we concluded that the present study does not allow determining accurately the margin proteome.

We further explored the proteins distribution among BBY and stroma-lamellae in order to reveal groups of proteins with similar abundance in BBY and stroma-lamellae. Thus, cluster analysis was performed on the BBY and stroma-lamellae relative mean abundances, calculated as the mean value across the 6 samples in each group, divided by the mean abundances across the thylakoid samples. The decision upon the number of clusters was taken by analyzing both the goodness of fit as well as the stability of a wide range of clustering solutions. By doing so, we ensured that the chosen solution would both find the data structures that best fit the actual data in terms of intergroup separability and cohesion, and would reveal the most stable data structures. Finally we partitioned the proteins into seven clusters, which are discussed in the following section. In addition, the clustering algorithm, its tuning and validation are thoroughly described in the Materials and Methods section.

In conclusion, an adapted statistical analysis showed that numerous proteins differentially distributed in either BBY or stroma-lamellae fractions could be identified whereas the margin fraction was characterized by a very low number of significantly distributed proteins.

Identification of Thylakoid Proteins that are Differentially Abundant between the BBY and the Stroma-lamellae Domains—As described above, differential analysis between BBY and stroma-lamellae samples allowed to point out 515 proteins, which were enriched either in BBY or stroma-lamellae fractions. However, a significant fraction represents contaminating proteins that were differentially found in the BBY and stroma-lamellae, because of differential contaminations of these two fractions by stroma or chloroplast envelope proteins. As discussed above, this leads to a mixed list of proteins that are genuinely enriched either in the BBY or in the stroma-lamellae fractions, along with stroma or chloroplast

envelope proteins that are differentially contaminating the purified thylakoid fractions.

In addition, exploring the data by cluster analysis provides a complementary view on the way that the proteins group based on their relative abundance in BBY and stroma-lamellae. Cluster analysis has been performed for exploratory reasons and has been conducted in parallel with differentially abundance analysis in order to observe whether we could identify, at first glance, additional groups of proteins known to have the same chloroplastic localization. The main gain added by cluster analysis is to validate the results obtained by differentially abundance analysis, in the sense that, many proteins identified as being differentially abundant in one compartment were also found as grouping together in a natural way by cluster analysis. The partition of proteins based on cluster analysis is available in [supplemental Table S1](#) and [supplemental Table S6](#), column PAM cluster/THY proteins. In addition the composition of clusters can be visualized in [supplemental Fig. S6](#). Thus, cluster 1 (203 proteins) contains proteins with a high relative abundance both in stroma-lamellae and BBY fractions, meaning that it contains the major BBY and stroma-lamellae proteins. As shown in Fig. 5 (black dots), the proteins within this cluster are spread both on the left and the right-hand side of the volcano-plot. Proteins of this cluster, that are members of the PSI and PSII photosystems, the cytochrome b_6f and the ATP synthase complexes will be particularly discussed (see below). On the contrary, the proteins grouped within cluster 2 (369 proteins) exhibit low relative abundance both in BBY and stroma-lamellae. Therefore, most of them could be considered minor proteins within these compartments. The proteins within clusters 3 (125 proteins), 4 (198 proteins), 5 (66 proteins) and 6 (31 proteins) appear to be stroma-lamellae specific proteins which present low relative abundance in BBY; the four clusters are separated according to their relative abundance in stroma-lamellae, the proteins in cluster 6 exhibiting the highest relative abundance, whereas the proteins in cluster 3, the lowest. Clusters 4 and 5 contain stroma-lamellae specific proteins with medium relative abundance. Note that in Fig. 5, the dots associated to the colors of these clusters are located on the right-hand side, which corresponds to proteins enriched in the stroma-lamellae fraction, as identified by the significance analysis of differentially abundance for BBY *versus* stroma-lamellae case. An interesting cluster is cluster 7 (48 proteins), which concentrates BBY-enriched proteins which were not, or hardly, detected in stroma-lamellae. Indeed in Fig. 5 (yellow dots), these proteins are located in the left-hand side, which corresponds to BBY proteins. Functional analysis of cluster 7 proteins shows that most of them are proteins related to chloroplast transcription (nucleoid-related proteins, DNA/RNA binding proteins, chloroplast transcriptionally active proteins or chloroplast-encoded RNA polymerase proteins subunits). All these proteins are annotated in the literature to be mainly located in the stroma. However the identification of nucleoid proteins in BBY

fractions is consistent with the fact that nucleoids are known to be located at the center of the chloroplast, in close vicinity to the thylakoids in mature chloroplasts (71, 72). In addition, Liu and Rose showed that nucleoids co-fractionate with the thylakoids and that a region of the chloroplast DNA is bound to the thylakoids (73). The present study shows that nucleoids are specifically and exclusively bound to BBY domains.

In order to be able to remove cross-contaminants deriving from stroma and envelope compartments, we performed a curated annotation of the 515 proteins that show a differential localization in the BBY or the stroma-lamellae fractions, to provide a detailed analysis of their subchloroplastic localization ([supplemental Table S6](#)). The protein names, their function or presence in a specific protein complex were carefully determined (see Materials and Methods section) as well as their subcellular and subchloroplastic localizations that were deduced from the screening of the literature, using previously known published information (e.g. from the AT_CHLORO and the PPDB databases) or from predictions using bioinformatics tools (e.g. ChloroP). After removing 12 proteins that are well-known major components of the plasma membrane, the tonoplast, the peroxisome or the mitochondria ([supplemental Table S6](#), ranks 504–515), proteins that were previously associated to the stroma or the envelope compartments ([supplemental Table S6](#), ranks 307–475), proteins from cluster 7 (except At2g34420; [supplemental Table S6](#), ranks 475–503) and additional stromal ribosomal proteins ([supplemental Table S6](#), ranks 295–306) 294 proteins were identified as being differentially distributed in BBY and stroma-lamellae fractions.

Thus, 27 and 267 thylakoid proteins were found to be more abundant in BBY and in stroma-lamellae fractions, respectively ([supplemental Table S6](#) and [supplemental Fig. S7](#)). These results both confirm current knowledge about localization of some classes of proteins and bring new insight over the protein content of BBY and stroma-lamellae, as discussed in the following sections.

An Unexpected Heterogeneity in the Subunit Distribution of the Photosynthetic Complexes—In order to get a better understanding on the differential subunits composition and distribution of the photosynthetic complexes in the fractions, we analyzed further the results obtained from differential analysis. Data obtained allowed us to corroborate the biochemical data concerning the differential localization of these complexes and to test possible differences in every complex between thylakoid fractions analyzed here. In particular, the LogFC value could be associated to the enrichment of a given protein in either BBY or stroma-lamellae fractions. We exploited this information to determine the relative subunit composition and abundance for four different photosynthetic maxi-complexes. Fig. 6 presents the results obtained in the case of the major components of the photosynthetic chain. Two different color gradations were used there to indicate enrichment in the BBY (brown) or in the stroma-lamellae (green) fractions respec-

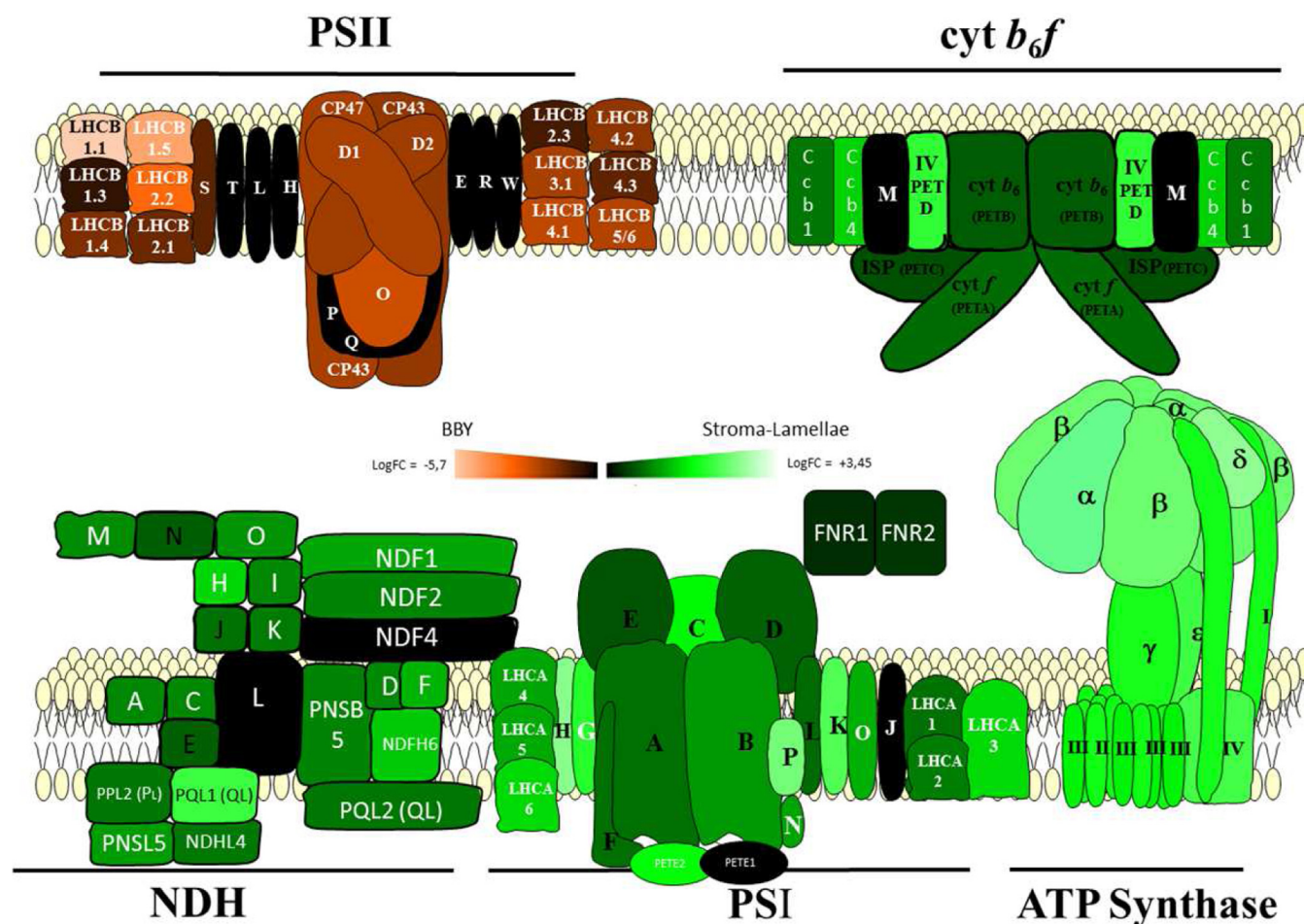


Fig. 6. Photosynthetic complexes distribution in the major thylakoid subfractions as determined by proteomic analysis. PSII, cytochrome *b₆f*, PSI, NDH and ATP synthase CF₀-F₁ complexes are represented. Schematic representations of the different photosynthetic complexes were redrawn according to Choquet and Vallon (110). Each subunit is indicated by its common name, or by the letter of the gene (*psb*, *pet*, and *psa*) that encodes it. Five color codes were employed to designate the specific localization of every subunit in the grana/stroma-lamella regions. Light brown indicates “high” preferential localization in the BBY ($\text{LogFC} \leq -1$). Dark brown indicates “medium” preferential localization in the BBY ($-1 < \text{LogFC} < 0$). Dark green indicates “medium” preferential localization in the stroma-lamellae ($0 < \text{LogFC} < 1$) whereas light green indicates “high” preferential localization in the stroma lamellae ($\text{LogFC} \geq 1$). Figs. were taken from [supplementary Table S6](#). The common name for proteins with a preferential localization can be found in [supplemental Table S6](#). ATPx, PSAs, NDHx, PTEs, and PSBs Subunits with no preferential localization were identified from the initial 1295 protein list (black color). Common names employed correspond to the following proteins: AtCg00630 (PsaJ), At1g70760 (NDHL), At3g16250 (NDF4 = CEF1), At2g26500 (PETM), At3g21055 (PsbT), AtCg00560 (PsbL), AtCg00710 (PsbH), AtCg00580 (PsbE), At1g79040 (PsbR), At4g28660 (PsbW), At1g06680 (PsbP), At4g21280, and At4g05180 (PsbQ).

tively for the four major photosynthetic complexes ([supplemental Table 6](#)): ATP synthase, cytochrome *b₆f*, PSII and PSI-NDH super complex. Subunits of the photosynthetic complexes that did not show differential localization (p value > 0.05) are marked in black.

The overall picture emerging from the present proteomic study is consistent with previous description of the localization of the main photosynthetic complexes in the thylakoid membranes (reviewed in (10, 65)): PSI and the ATP synthase complexes are accumulated in the unstacked stroma-lamellae (together with the chlororespiratory complex NDH), whereas PSII is mostly found in the BBY stacks (Fig. 6). The cytochrome *b₆f* complex is more ubiquitous, although slightly more concen-

trated in the stroma-lamellae. However, a closer look at these data reveals an unexpected heterogeneity in the subunit composition of the different complexes. In particular within a given complex, the localization of different subunits between the BBY/stroma-lamellae turned out to be variable, as clearly evidenced by the case of PSI. As expected, this complex accumulates preferentially in the stroma-lamellae fraction. However, not all its subunits show the same ratio between the stroma-lamellae and the BBY fractions. For example the PsaC subunit had a higher score for stroma-lamellae than other nearby core subunits (e.g. PsaD). This finding was confirmed by Western blot analysis using specific antibodies ([supplemental Fig. S8](#)), showing that the amount of PsaC found in the BBY samples is lower than in

the case of PsaD. In principle, this finding could reflect a purification artifact. Some peripheral PSI subunits (e.g. PsaK, PsaG, and PsaH), which also share with PsaC the property of being enriched in stroma lamellae, could have been differentially extracted by the two types of detergents used to isolate the BBY and the stroma-lamellae, leading to an apparent enrichment in a specific fraction. To test this hypothesis, we performed experiments to analyze the sensitivity of these proteins toward extraction by detergents. Thylakoids were incubated with increasing concentration of digitonin or Triton and the amount of proteins present in the supernatant and the pellet was compared. We focused on three PSI subunits: an intrinsic one (PsaB) and two more peripheral ones (PsaC, PsaD) and found that they were all similarly extracted by both types of detergents. Thus it appears that the differential enrichment of some PSI subunits in the BBY and stroma lamellae fractions is not caused by a different sensitivity to detergent, suggesting instead that the PSI subunit composition may vary depending on its location in the thylakoids.

DISCUSSION

Distribution of the Photosynthetic Complexes—This article addresses the protein composition in the main subcompartments of the thylakoid membranes: the grana (here represented by the innermost part, the BBY) and the stroma-lamellae. As discussed above, our differential analysis allows discriminating these two fractions, while does not allow determining accurately the proteome of the third thylakoid compartment: the margins. This differential analysis is extremely informative to reveal possible differences in the composition of super-complexes associated either to stroma-lamellae or BBY membranes. Furthermore, it may unravel functional differences between complexes localized in the two fractions. In most cases, a differential and specific repartition between stroma-lamellae and BBY was found. The ATP synthase CF_0 - F_1 is the most homogeneous complex according to our analysis. All its subunits are equally (and almost exclusively) found in the stroma-lamellae with a strong level of enrichment ($\text{LogFC} > 2$) (Fig. 6). PSI is also enriched in this fraction although to a lesser extent. If we refer to the LogFC value, the finding that PSI is more present in the BBY fraction (relatively to occurrence in the stroma-lamellae fraction) than the ATP synthase CF_0 - F_1 is not surprising, as it corroborates previous ideas on the necessity for this complex to accumulate in the PSII rich regions to ensure proper electron flow (65). As discussed above, this complex shows a larger heterogeneity at the level of its subunits, which according to our data, does not reflect a purification artifact but rather a genuine difference in the subunit composition between complexes present in the stroma-lamellae and the BBY particles. The finding that PsaC (AtCg01060) is more represented than other subunits in the stroma-lamellae fractions is surprising, as this subunit has been previously shown to represent the docking site for the interaction of the PsaD and PsaE ones during complex as-

sembly in thylakoids (74). Indeed previous analysis in *Chlamydomonas* has revealed that mutants lacking this subunit cannot assemble the PSI complex (75). Moreover, this subunit bears two of the three terminal FeS centers, suggesting that PSI complexes devoid of this subunit should not be functional. However, previous data obtained from cyanobacteria, suggests that PSI can accumulate in a functional state even in the absence of this PsaC subunit (76). This opens the possibility that the PSI complexes found in the BBY fraction could represent a genuine form of this complex, still bearing some electron flow capacity. Other peripheral PSI subunits, PsaG (At1g55670), PsaH (At3g16140, At1g52230), and PsaK (At1g30380) also show a preferential location in the stroma-lamellae. Among them, PsaG and K are involved in the docking of the PSI antennas (LHCI) binding to the core complex (66). It is interesting to note that Lhca3 (At1g61520), which is directly bound to PsaK in the 3D structure of this complex (66), is also preferentially localized in the stroma-lamellae. It is tempting therefore to propose that this PsaK-Lhca3 subcomplex could be partially released from the PSI complex present in the grana. Another subunit, PsaH, is required for the docking of LHCII to PSI during state transitions (77), that is, the redox induced reversible phosphorylation of PSII antenna, which lead to their displacement from PSII and binding to PSI (78). Thus, the finding that this subunit is enriched in the stroma-lamellae PSI complex is consistent with the occurrence of LHCII docking to PSI in the stromal fraction during state transitions. The hypothesis of a different subunit composition in the BBY and stroma-lamellae PSI fractions is also plausible based on the finding that the two minor LHCI species (Lhca5, At1g45474 and Lhca6, At1g19150) were also highly enriched in the stroma-lamellae. These subunits provide a molecular platform for the interaction between PSI and the NDH complex, leading to the formation of a PSI-NDH super-complex, which is found in the stroma-lamellae (79). In our proteomic survey, most of the subunits previously attributed to this super-complex co-accumulate in the stroma-lamellae, suggesting a rather homogeneous structure of this super-complex. Moreover, we noticed that, besides PPL2 (At2g39470) and the two PsbQ-like proteins (At1g14150 and At3g01440) previously found in this complex, several other pseudo-PSII subunits specifically localized in the stroma-lamellae. It is tempting to propose that these PsbP-like subunits, that show a totally different localization when compared with PSII, could also be linked to the NDH complex. The finding of a PSI-NDH super-complex in the stroma-lamellae is corroborating the notion that this chlororespiratory complex is required for cyclic electron flow around PSI (80). Consistent with this, we found that the Pgr5 (At2g05620) and Pgr1 isoforms (At4g11960, At4g22890), which are also required for cyclic electron flow in plants (80) were also highly enriched in the stroma-lamellae fractions. Based on these results, it is therefore tempting to speculate that the heterogeneous composition of PSI complexes seen in the two compartments may

reflect a different functional role of PSI in the two compartments: PSI in the grana has been proposed to participate in linear electron flow working in series with PSII, whereas stromal PSIs would mainly perform cyclic electron flow.

Among the major photosynthetic complexes, the cytochrome b_6f is the one showing the most homogeneously localization in the thylakoids, being only moderately concentrated in the stroma-lamellae. Overall, the fraction of the cytochrome b_6f that is localized in the BBY appears to be higher in this analysis than in previous reports based on immunolocalization (81, 82). However, our findings are consistent with previous investigations on thylakoid fractions isolated upon mechanical fractionation of these membranes, followed by biochemical analysis (e.g. (83) and references therein). Indeed, most of its core subunits (cytochrome b_6 , AtCg00720, and f , AtCg00540, and the Rieske iron-sulfur protein PetC, At4g03280) showed similar relative enrichment. This is not the case for subunit IV (AtCg00730), which is a central subunit of this complex and is most exclusively located in the stroma-lamellae. This is again unexpected, as cytochromes b_6f complexes should not be able to accumulate in the absence of subunit IV (84). At present no explanation can be proposed for this finding. Besides subunit IV, the CCB proteins (CCB1, At3g26710, and CCB4, At1g59840), that is, the chaperones required for the assembly of the high spin heme c' to the stromal side of the complex (85), are also largely enriched in the stroma-lamellae fractions. These subunits however are not required for complex activity but only for complex assembly. Thus their peculiar localization can be rationalized based on earlier suggestions that the unstacked membranes are the site where photosynthetic complexes are assembled or repaired, as previously demonstrated in the case of the insertion of the subunit D1 (PsbA, AtCg00020) into PSII during the repair cycle which follows photoinhibition (86). If the same scenario is translated to the cytochrome b_6f complex, it is reasonable to envisage that insertion of the c' heme, which requires rather complex molecular machinery, would also preferentially take place in membranes that are easily accessible from the stroma, i.e. the stroma-lamellae. The conclusion that assembly/repair of the photosynthetic complexes is located in the stroma-lamellae is expected (86) and also supported by the finding that all the thylakoid-bound proteases (Deg, FtsH and Clp), which are required for disassembly of these complexes, are enriched in the stroma-lamellae (see next section).

As expected, the only photosynthetic complex that is preferentially located in the BBY fraction is PSII. Again, a significant heterogeneity is seen at the level of its different subunits. Although all the core complex subunits including D1 (PsbA), D2 (PsbD, AtCg00270), CP43 (PsbC, AtCg00280), CP47 (PsbB, AtCg00680), and PsbO (At5g66570, At3g50820) are almost exclusively found in the BBY, other subunits of the water-oxidizing complex (PsbP and PsbQ) have a less defined localization. This difference could, in principle, reflect a problem in distinguishing

between the true PSII subunits (PsbP1 At1g06680, PsbQ1 At4g21280, and PsbQ2 At4g05180) and the so called PsbP-like and PsbQ-like proteins, which are not bound to PSII but rather located in the stroma-lamellae, being possibly bound to the NDH complex (see above; PQL1, At1g14150, and PQL2, At3g01440, proteins). Note that PsbP-like were all found in cluster 3, which reflects similar relative abundance and behavior. PsbS (At1g44575), the small PSII subunit involved in photoprotection via the induction of enhanced thermal dissipation in the PSII antenna (87) also showed a less pronounced accumulation in the grana stack than the core complex subunits. This finding is consistent with earlier investigation (88) suggesting that this protein provides a flexible link between the PSII core and its antenna complexes. Therefore, this protein could be easily detached from the core complex during sample purification. However, a possible localization of some PsbS in the non appressed regions cannot be totally excluded based on recent data in *Chlamydomonas*, where a displacement of LHCSR3 (the functional homolog of PsbS in microalgae) between PSII and PSI was observed depending on the physiological conditions (89). Finally, the distribution of the minor (Lhcb 4 At5g01530, At3g08940, At2g40100, Lhcb 5 At4g10340, and Lhcb 6 At1g15820) and trimeric (Lhcb 1 At1g29920, At1g29930, At2g34430, At2g34420, Lhcb 2 At2g05100, At2g05070, At3g27690, and Lhcb 3 At5g54270) light harvesting complexes between the stroma-lamellae and BBY fractions is much more heterogeneous than that of the core PSII subunits. In principle, this could reflect the existence of antenna complexes with a different mobility in the membranes (e.g. the L, M and S LHCI complexes, (90)) which could differentially move to the stroma-lamellae during the state 1 to state 2 transition (78). However, our data concerning the differential location of PSII antenna in the two fractions are not consistent with the recent detailed analysis of the mobility of the different LHCI subunits during state transitions (91). This suggests that other structural causes are probably responsible for the observed heterogeneity. As an alternative hypothesis, we propose therefore that the peculiar distribution of LHCI could reflect their migration to the stroma-lamellae during the PSII reparation cycle that follows photoinhibition (86).

Eventually, the finding that a nonnegligible fraction of the two FNR (ferredoxin-NADP reductase, At5g66190 and At1g20020) isoforms is located in the BBY is unexpected based on previous results, which indicates that FNR should be completely absent from this compartment (10). In principle this could be explained by the finding that cyt b_6f complexes is rather abundant in our BBY fractions, at variance with previous reports (10). Because FNR is bound to this complex (92) its presence could simply reflect the cyt b_6f enrichment in BBY.

Thylakoid FTSH and DEG Proteases—FtsH proteins are membrane-bound ATP-dependent metalloproteases. FtsH proteases localized in the chloroplast have been shown to

play a major role in assembly and maintenance of the plastid membrane system (for review, see (93)). In *Arabidopsis*, only Ftsh1 (At1g50250), 2 (At2g30950), 5 (At5g42270), and 8 (At1g06430) have been shown to reside in the thylakoid membranes. In this study, those four proteases were identified in both thylakoid subcompartments, with a strong enrichment in the stroma-lamellae. This is in good agreement with recent data showing that Ftsh proteins were present in both thylakoid subcompartments, as monomers and homo/heteromeric dimers in stromal region and as hexameric complexes in the grana regions (94). The thylakoid located Ftsh complex in *Arabidopsis* is responsible for degradation of photodamaged D1 protein (PsbA) in concert with luminal Deg proteases (see for reviews (95, 96)). All Ftsh proteins that were found differentially located in the stroma-lamellae fractions belong to cluster 1, indicating that they are abundant.

Deg proteins are serine proteases and, unlike Ftsh, are not ATP-dependent. It is assumed that in plants Ftsh digests D1 after the first proteolytic cleavage has been performed by a Deg protease (96). The Deg1 (At3g27925), Deg5 (At4g18370) and Deg8 (At5g39830) were previously reported to be present in the thylakoid lumen. In the present study, only Deg1 was identified as being significantly enriched in the stroma-lamellae. This protease has been shown to be involved in D1 protein degradation but also in PSII assembly (97). Thus one could expect that Deg1 could be present in the stroma lamellae, there being involved in the PSII repair cycle (86). Four other Deg proteases were identified in our 1295 list of proteins (Deg2 At2g47940, Deg3 At1g65630, Deg5, and Deg8). Except Deg1, most of these Deg proteins were found in cluster 2, which gathers low abundant proteins. In addition Deg2, 3, 5, and 8 were not significantly found to be enriched in either the BBY fraction or the stroma-lamellae fractions. However Deg5 was identified in BBY fractions but not in stroma-lamellae, which indicates that Deg5 might be enriched in BBY. This observation is consistent with a role for Deg5 in the repair of damaged PSII, possibly by performing an initial cleavage of the D1 protein within lumen-directed loops (98). Also, Deg2 and Deg3 were detected in stroma-lamellae but not in BBY fractions. Thus Deg2 is likely to be more enriched in stroma-lamellae, which is consistent with the fact that Deg2 was found to be peripherally attached to the stromal side of the thylakoid membrane. This protease could be part of a large network of enzymes that ensure protein quality control in PSII, and could be also involved in the degradation of Lhcb6, the minor light-harvesting protein of PSII (99).

Transporters of the Thylakoid Sub-compartments—Few thylakoid transporters have been identified in the present study, probably because samples were analyzed without specific enrichment of hydrophobic proteins before mass spectrometry analysis. Indeed, transporters are often minor and highly hydrophobic proteins, and their identification needs dedicated treatments like solubilization in organic solvent or in detergent (100), prior to separation on SDS-PAGE and trypsin

digestion. About ten transporters were shown to be differentially distributed between the stroma-lamellae and the BBY substructures. The only transporter known to be located in the thylakoids and identified as being differentially distributed is AtHMA8 (At5g21930), a copper transporter belonging to the P_{1B} -type ATPases family. This transporter allows the import of copper into the lumen of thylakoids to supply the plastocyanin which uses copper as cofactor. From characterization of *Arabidopsis hma8* mutant and investigation of the localization of AtHMA8 by several approaches (transient expression of GFP fusion and *in vitro* import experiments on a truncated precursor), it was suggested that this transporter was localized in the thylakoids (68). In the present work, AtHMA8 was identified in stroma-lamellae but not in BBY fractions. This is consistent with the localization of the two isoforms of plastocyanin (Pete2, At1g20340, and Pete1, At1g76100) that were found both in BBY and stroma-lamellae fractions for the Pete1 isoform or highly enriched in the stroma-lamellae for the Pete2 isoform. Four putative thylakoid transporters NTF2 (At1g71480), KEA3 (At4g04850), NHD1 (At3g19490), and ABCD2 (At1g54350) were also found to be differentially distributed and were previously identified in chloroplasts (15) or plasma membrane for NHD1 (101). In this study, NTF2, KEA3, NDH1 and ABCD2 were all detected in the stroma-lamellae but none were identified in BBY fractions, suggesting that they should be involved in transport of ions or metabolic specifically required in stroma-lamellae. Six additional transporters were identified, like the phosphate-triose phosphate (TPT, At5g46110) and the oxoglutarate/malate translocators (DiT1, At5g12860, and DiT2-1, At5g64290), that were already known to be associated with the chloroplast envelope. In the AT_CHLORO database, they have been associated mainly with the envelope but also with the thylakoid membranes. The phosphate and triose phosphate and oxoglutarate and malate translocators with the protein OEP16-1 (At2g28900) were all found in BBY and stroma-lamellae fractions with enrichment in the latter sub-fraction. This specific enrichment is consistent with a contamination by envelope membranes which is higher in stroma-lamellae than in BBY fraction, the envelope membrane having similar density compared with the stroma-lamellae's one. The last three transporters, IEP18 (At5g62720), BASS2 (At2g26900), and NHD1 (At3g19490) were not identified in the BBY fraction. Because, the function of IEP18 is still unknown, we cannot conclude about the relevance of its localization both in chloroplast envelope and in the stroma-lamellae. Characterization of an *Arabidopsis bass2* mutant suggested that this protein is a sodium-dependent pyruvate transporter of the chloroplast envelope (102). In this study, the authors also suggested that sodium influx could be balanced by the sodium:proton antiporter NHD1. This transporter, previously identified both in chloroplast and plasma membrane proteomes (15, 101), was recently characterized as a sodium exporter of the chloroplast envelope (103). In the AT_CHLORO database, BASS2 was

found associated with both envelope and thylakoid membranes, although NDH1 was not detected. Thus, beside their characterized function as transporters of the chloroplast envelope, we cannot exclude a dual localization in subchloroplast compartment with a physiological function that remains to be determined in the thylakoids.

One question that we wanted to address through identification of the stroma-lamellae proteome concerns the controversy about envelope localization of some transporters. Indeed, both the ATP/ADP carrier TaaC (At5g01500) and Na⁺-dependent phosphate transporter Pht4;1 (ANTR1, At2g29650) were found exclusively associated to the envelope in the AT_CHLORO database (14) although they were described as thylakoids transporters (104, 105). Moreover, none of these two proteins were identified in proteomic investigations dedicated to the study of membrane or lumen thylakoid (supplemental Table S4). In the present study, both proteins were found to be enriched in the stroma-lamellae fraction. Thus, considering previous proteomic studies and present enrichment in stroma-lamellae fractions, TaaC and Pht4;1 were considered as being likely contaminants from the envelope. The envelope localization was confirmed by Western-blot analyses in which the phosphate transporter Pht4;1 was only detected in envelope fraction and no signal appears in the stroma-lamellae (supplemental Fig. S9). Furthermore, a recent study showed that the TAAC transporter is indeed a 3'-phosphoadenosine 5'-phosphosulfate/5'-phosphoadenosine 3'-phosphate transporter of the chloroplast envelope (106), in good agreement with a chloroplast envelope localization and thus our proteomic data ((14) and this work).

Potential New Thylakoid Proteins—From the present study we could identify 218 proteins that were not previously identified in proteomic studies dedicated to the chloroplast or the thylakoids (supplemental Table S5). At least 103 of them might be genuine thylakoid (or at least chloroplast) proteins but very few of the remaining 115 proteins could be further considered for tentative chloroplast localization because most of them are either totally unknown proteins or only contain domains that allow to suspect or predict a putative function (supplemental Table S5, rank 104–218). However, quite surprisingly, components of the ubiquitin-proteasome system were detected among those 115 proteins. A recent study has demonstrated that plastid biogenesis is directly regulated by the ubiquitin-proteasome system (107). Using forward genetics, a new outer chloroplast envelope protein was identified, SP1 (At1g63900) which encodes a RING-type ubiquitin E3 ligase that mediates ubiquitination of the TOC translocon. Interestingly, while SP1 was not detected during the present study (as expected because purified envelope fractions were not analyzed during this work), several proteins could be identified that share some functional similarities with SP1. Indeed, one protein of the F-box/kelch-repeat superfamily (At4g39590) and a protein of the RING/U-box superfamily (At5g40140) are present in the list of 103 proteins that are predicted to be targeted to the chloroplast using ChloroP and

TargetP. Furthermore, nine other components of the ubiquitin-proteasome system (At1g22500, At3g59250, At5g03100, At1g06900, At5g64760, At5g60250, At2g24540, At5g03100, At1g78100, At5g05560) were detected in the analyzed chloroplast fractions and are listed in the list of 115 proteins that are not predicted to be targeted to the chloroplast. Although we have no other explanation than cross-contamination events for the detection of these last nine components of the ubiquitin-proteasome system in the various thylakoid sub-compartments that were analyzed during this study (BBY, stroma-lamellae or margins), identification of these proteins at least suggests that present-days proteomic studies reveal very minor components that were not accessible during previous analyses.

In the list of 103 potentially new thylakoid proteins, we performed an exhaustive analysis of the known or predicted functions of these previously undetected chloroplast or thylakoid components. As expected, almost one third of these proteins have no known function or even predicted functional domain (supplemental Table S5 see column “Curated function (this work)”). Around 10% of these newly identified proteins might be involved in lipid metabolism or transport (lipases, desaturases, acyl-transferases etc.). Seven proteins are known or predicted to be involved in the metabolism of vitamins or pigments and four putative ion or metabolite transporters could also be detected. Although 15 different LhcII isoforms were previously detected in large-scale proteomic analyses, including an LhcII2.3 protein (see supplemental Table S1), an LhcII2.3-like protein (At1g76570) is listed in the short list of four proteins that are directly linked to photosynthesis. More interestingly, this list of 103 proteins also contains numerous regulatory proteins like the two proteins that are involved in cyclic electron flow around the Photosystem I, that is, CEF1 (At3g16250) and NDHL (At1g70760) (108), or the cytochrome *b6f* biogenesis protein CCDA (At5g54290) (109) that could not be detected in previous proteomic studies targeted to the chloroplast or the thylakoid membranes. Beside proteins that might be additional actors of the regulation of the chloroplast physiology, eleven detected proteins are suspected to be chaperones or proteases, four proteins are predicted kinases, one is a putative kinase regulator, one is a phosphatase, and seven are putative RNA binding proteins (mostly TPR or PPR proteins).

CONCLUSION

The investigation presented here targets the thylakoid membrane of the chloroplast that harbors the light-dependent reactions of photosynthesis by which green plants synthesize organic compounds from water and carbon dioxide. This work, performed on the model organism *Arabidopsis*, provides a strong basis to carry out targeted approaches in physiologically and economically important organisms such as crops or green algae.

The present work is the first study designed to address the precise localization of thylakoid proteins with respect to two major subcompartments: the grana and the stroma-lamellae, using a proteomics-based approach. The present work provides a more exhaustive repertoire of minor chloroplast proteins, thus representing a new resource for thylakoid proteins with respect to both functional and localization issues. Thus, this study complements our previous large-scale approach aiming to provide detailed information about the subchloroplast localization of proteins (14).

Overall, the data from the proteomic study presented here corroborates previous conclusions on the distribution of photosynthetic complexes in the thylakoids. However, thanks to an in depth proteomic analyses coupled to a dedicated data analysis pipeline, new information with respect to several aspects of the thylakoid proteome has been achieved. Indeed, the present data unravels a structural heterogeneity between the photosynthetic complexes located in the different compartments, which is particularly evident in the case of the two photosystems. Differential protein segregation of PSII or PSI proteins in the thylakoid compartments may reflect the different content of antenna complexes that was previously seen between stromal and granal PSII complexes (9). This segregation also strongly suggests that PSI needs to recruit new subunits to allow the formation of the PSI-NDH supercomplex in the stroma-lamellae. Now that we have achieved a robust map of proteomes within thylakoid membranes, then next step will be to carry out analysis of the functional consequences of the structural heterogeneity of photosynthetic complexes by employing targeted experiments to screen the impact of knock-out mutations or of different environmental conditions on the regulatory processes that control these protein segregations.

* This study received financial support from the French National Research Agency (ANR-2010-GENOM-BTV-002-01 Chloro-Types; ANR-10-LABEX-04 GRAL Labex, Grenoble Alliance for Integrated Structural Cell Biology; ANR-10-INBS-08 ProFI (Proteomics French Infrastructure) and from EU FP7 program (Prime-XS project, Contract no. 262067) and the Marie Curie Initial Training Network Accliphot (FP7-PEOPLE-2012-ITN, grant agreement number 316427).

[S] This article contains supplemental Figs. S1 to S9 and Tables S1 to S7.

||| To whom correspondence should be addressed: Univ. Grenoble Alpes, 17 rue des Martyrs, F-38000 Grenoble, 38054 France. Tel.: 33-04-04-38-78-25-49; E-mail: myriam.ferro@cea.fr; norbert.rolland@cea.fr.

REFERENCES

1. Rolland, N., Curien, G., Finazzi, G., Kuntz, M., Maréchal, E., Matinge, M., Ravel, S., and Seigneurin-Berny, D. (2012) The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annu. Rev. Genet.* **46**, 233–264
2. Joyard, J., Ferro, M., Masselon, C., Seigneurin-Berny, D., Salvi, D., Garin, J., and Rolland, N. (2010) Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. *Prog. Lipid Res.* **49**, 128–158
3. Joyard, J., Ferro, M., Masselon, C., Seigneurin-Berny, D., Salvi, D., Garin, J., and Rolland, N. (2009) Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. *Mol. Plant* **2**, 1154–1180
4. Block, M. A., Douce, R., Joyard, J., and Rolland, N. (2007) Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. *Photosynth. Res.* **92**, 225–244
5. Weber, A. P., Schwacke, R., and Flügge, U. I. (2005) Solute transporters of the plastid envelope membrane. *Annu. Rev. Plant Biol.* **56**, 133–164
6. Joyard, J., Teyssier, E., Mieg, C., Berny-Seigneurin, D., Marechal, E., Block, M. A., Dorne, A. J., Rolland, N., Ajlani, G., and Douce, R. (1998) The biochemical machinery of plastid envelope membranes. *Plant Physiol.* **118**, 715–723
7. Tonkin, C. J., Kalanion, M., and McFadden, G. I. (2008) Protein targeting to the malaria parasite plastid. *Traffic* **9**, 166–175
8. Oborník, M., Janouskovec, J., Chrudimský, T., and Lukes, J. (2009) Evolution of the apicoplast and its hosts: from heterotrophy to autotrophy and back again. *Int. J. Parasitol.* **39**, 1–12
9. Rochaix, J. D. (2013) Redox regulation of thylakoid protein kinases and photosynthetic gene expression. *Antioxid. Redox. Signal.* **18**, 2184–2201
10. Dekker, J. P., and Boekema, E. J. (2005) Supramolecular organization of thylakoid membrane proteins in green plants. *Biochim. Biophys. Acta* **1706**, 12–39
11. Agarwal, R., Matros, A., Melzer, M., Mock, H. P., and Sainis, J. K. (2010) Heterogeneity in thylakoid membrane proteome of *Synechocystis* 6803. *J. Proteomics* **73**, 976–991
12. Lilley, K. S., and Dupree, P. (2007) Plant organelle proteomics. *Curr. Opin. Plant Biol.* **10**, 594–599
13. Dunkley, T. P., Hester, S., Shadforth, I. P., Runions, J., Weimar, T., Hanton, S. L., Griffin, J. L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R. B., Dupree, P., and Lilley, K. S. (2006) Mapping the *Arabidopsis* organelle proteome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6518–6523
14. Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Le Gall, S., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., and Rolland, N. (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell. Proteomics* **9**, 1063–1084
15. Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and van Wijk, K. J. (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **3**, e1994
16. Kleffmann, T., Hirsch-Hoffmann, M., Grisse, W., and Baginsky, S. (2006) p1prot: a comprehensive proteome database for different plastid types. *Plant Cell Physiol.* **47**, 432–436
17. Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjölander, K., Grisse, W., and Baginsky, S. (2004) The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14**, 354–362
18. Kieselbach, T., Bystedt, M., Hynds, P., Robinson, C., and Schröder, W. P. (2000) A peroxidase homologue and novel plastocyanin located by proteomics to the *Arabidopsis* chloroplast thylakoid lumen. *FEBS Lett.* **480**, 271–276
19. Kieselbach, T., Hagman, Andersson, B., and Schröder, W. P. (1998) The thylakoid lumen of chloroplasts. Isolation and characterization. *J. Biol. Chem.* **273**, 6710–6716
20. Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adam-ska, 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**, 319–341
21. Peltier, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Söderberg, 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
22. Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schröder, W. P., and Kieselbach, T. (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* **277**, 8354–8365
23. Goulas, E., Schubert, M., Kieselbach, T., Kleczkowski, L. A., Gardestrom, P., Schröder, W., and Hurry, V. (2006) The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short- and long-term exposure to low temperature. *Plant J.* **47**, 720–734

24. van Wijk, K. J. (2004) Plastid proteomics. *Plant Physiol. Biochem.* **42**, 963–977
25. Kieselbach, T., and Schröder, W. P. (2003) The proteome of the chloroplast lumen of higher plants. *Photosynth. Res.* **78**, 249–264
26. Whitelegge, J. P., Gundersen, C. B., and Faull, K. F. (1998) Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci.* **7**, 1423–1430
27. Vener, A. V., Harms, A., Sussman, M. R., and Vierstra, R. D. (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *J. Biol. Chem.* **276**, 6959–6966
28. Zolla, L., Timperio, A., Walcher, W., and Huber, C. G. (2003) Proteomics of light-harvesting proteins in different plant species. Analysis and comparison by liquid chromatography-electrospray ionization mass spectrometry. *Photosystem II. Plant Physiol.* **131**, 198–214
29. Zolla, L., Rinalducci, S., Timperio, A. M., and Huber, C. G. (2002) Proteomics of light-harvesting proteins in different plant species. Analysis and comparison by liquid chromatography-electrospray ionization mass spectrometry. *Photosystem I. Plant Physiol.* **130**, 1938–1950
30. Gómez, 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
31. Gómez, S. M., Bil', K. Y., Aguilera, R., Nishio, J. N., Faull, K. F., and Whitelegge, J. P. (2003) Transit peptide cleavage sites of integral thylakoid membrane proteins. *Mol. Cell. Proteomics* **2**, 1068–1085
32. Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., and Wijk, K. J. (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* **16**, 478–499
33. Peltier, J. B., Ripoll, D. R., Friso, G., Rudella, A., Cai, Y., Ytterberg, 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 three-dimensional structures, and functional implications. *J. Biol. Chem.* **279**, 4768–4781
34. Turkina, M. V., Kargul, J., Blanco-Rivero, A., Villarejo, A., Barber, J., and Vener, A. V. (2006) Environmentally modulated phosphoproteome of photosynthetic membranes in the green alga *Chlamydomonas reinhardtii*. *Mol. Cell. Proteomics* **5**, 1412–1425
35. Zolla, L., Rinalducci, S., and Timperio, A. M. (2007) Proteomic analysis of photosystem I components from different plant species. *Proteomics* **7**, 1866–1876
36. Salvi, D., Rolland, N., Joyard, J., and Ferro, M. (2008) Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. *Methods Mol. Biol.* **432**, 19–36
37. Porra, R. J., Thomson, 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
38. Berthold, D. A., Babcock, G. T., and Yocum, C. F. (1981) A highly resolved, oxygen-evolving photosystem II preparation from spinach thylakoid membranes. *FEBS Lett.* **134**, 231–234
39. Chua, N. (1980) Electrophoretic analysis of chloroplast proteins. *Methods Enzymol.* **69**, 434–466
40. Seigneurin-Berny, D., Gravot, A., Auroy, P., Mazard, C., Kraut, A., Finazzi, G., Grunwald, D., Rappaport, F., Vavasseur, A., Joyard, J., Richaud, P., and Rolland, N. (2006) HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *J. Biol. Chem.* **281**, 2882–2892
41. Ferro, M., Salvi, D., Brugière, 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
42. Dupieris, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., and Bruley, C. (2009) A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics* **25**, 1980–1981
43. Fermin, D., Basur, V., Yocum, A. K., and Nesvizhskii, A. I. (2011) Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **11**, 1340–1345
44. R Development Core Team (2008). R: a language and environment for statistical computing.
45. Gatto, L., and Lilley, K. S. (2012) MSnbase—an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **28**, 288–289
46. Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T., and Lilley, K. S. (2014) Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinformatics* **30**, 1322–1324
47. Karpievitch, Y. V., Dabney, A. R., and Smith, R. D. (2012) Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13**, S5
48. Little, R. J. A., and Rubin, D. B. (2002). Statistical analysis with missing data. Edition (Wiley-Interscience).
49. Luo, R., Colangelo, C. M., Sessa, W. C., and Zhao, H. (2009) Bayesian analysis of iTRAQ data with nonrandom missingness: identification of differentially expressed proteins. *Stat. Biosci.* **1**, 228–245
50. Karpievitch, Y., Stanley, J., Taverner, T., Huang, J., Adkins, J. N., Ansong, C., Heffron, F., Metz, T. O., Qian, W. J., Yoon, H., Smith, R. D., and Dabney, A. R. (2009) A statistical framework for protein quantitation in bottom-up MS-based proteomics. *Bioinformatics* **25**, 2028–2034
51. Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520–525
52. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193
53. Smyth, G. K. (2005). Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, 397–420.
54. Kooperberg, C., Aragaki, A., Strand, A. D., and Olson, J. M. (2005) Significance testing for small microarray experiments. *Stat. Med.* **24**, 2281–2298
55. Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300
56. Park, H., and Jun, C. (2009) A simple and fast algorithm for k-medoids clustering. *Expert Sys. Appl.* **36**, 3336–3341
57. Dhillon, I. S., Guan, Y., and Kulis, B. (2004). Kernel k-means: spectral clustering and normalized cuts. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 551–556
58. Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab—an S4 package for kernel methods in R. Report.
59. Courty, N., Burger, T., and Marteau, P. (2012). Geodesic analysis on the gaussian RKHS hypersphere. *Lect. Notes Artif. Int.* 299–313
60. Sun, Q., Zybaïlov, B., Majeran, W., Friso, G., Olinares, P. D., and van Wijk, K. J. (2009) PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Res.* **37**, D969–74
61. Tanz, S. K., Castleden, I., Hooper, C. M., Vacher, M., Small, I., and Millar, H. A. (2013) SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res.* **41**, D1185–91
62. Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939
63. Klie, S., and Nikoloski, Z. (2012) The choice between mapman and gene ontology for automated gene function prediction in plant science. *Front. Genet.* **3**, 115
64. Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L. (2010) Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**, 2272–2281
65. Albertsson, P. (2001) A quantitative model of the domain structure of the photosynthetic membrane. *Trends Plant Sci.* **6**, 349–358
66. Nelson, N., and Yocum, C. F. (2006) Structure and function of photosystems I and II. *Annu. Rev. Plant Biol.* **57**, 521–565
67. Glynn, J. M., Froehlich, J. E., and Osteryoung, K. W. (2008) *Arabidopsis* ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *Plant Cell* **20**, 2460–2470

68. Abdel-Ghany, S. E., Müller-Moulé, P., Niyogi, K. K., Pilon, M., and Shikanai, T. (2005) Two P-type ATPases are required for copper delivery in *Arabidopsis thaliana* chloroplasts. *Plant Cell* **17**, 1233–1251
69. Lazar, C., Taminiau, J., Meganck, S., Steenhoff, D., Coletta, A., Molter, C., de Schaetzen, V., Duque, R., Bersini, H., and Now, A. (2012) A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Transactions on Comput. Biol. Bioinform.* **9**, 1106–1119
70. Mamedov, F., Stefansson, H., Albertsson, P. A., and Styring, S. (2000) Photosystem II in different parts of the thylakoid membrane: a functional comparison between different domains. *Biochemistry* **39**, 10478–10486
71. Sato, N., Rolland, N., Block, M. A., and Joyard, J. (1999) Do plastid envelope membranes play a role in the expression of the plastid genome? *Biochimie (Paris)* **81**, 619–629
72. Sato, N., Albrieux, C., Joyard, J., Douce, R., and Kuroiwa, T. (1993) Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids. *EMBO J.* **12**, 555–561
73. Liu, J. W., and Rose, R. J. (1992) The spinach chloroplast chromosome is bound to the thylakoid membrane in the region of the inverted repeat. *Biochem. Biophys. Res. Commun.* **184**, 993–1000
74. Mannan, R. M., Pakrasi, H. B., and Sonoike, K. (1994) The Psac protein is necessary for the stable association of the Psad, PsaE, and PsaL proteins in the photosystem I complex: analysis of a cyanobacterial mutant strain. *Arch. Biochem. Biophys.* **315**, 68–73
75. Takahashi, Y., Goldschmidt-Clermont, M., Soen, S. Y., Franzén, L. G., and Rochaix, J. D. (1991) Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertional inactivation of the psac gene encoding the iron sulfur protein destabilizes photosystem I. *EMBO J.* **10**, 2033–2040
76. Mannan, R. M., Whitmarsh, J., Nyman, P., and Pakrasi, H. B. (1991) Directed mutagenesis of an iron-sulfur protein of the photosystem I complex in the filamentous cyanobacterium *Anabaena variabilis* ATCC 29413. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10168–10172
77. Lunde, C., Jensen, P. E., Haldrup, A., Knoetzel, J., and Scheller, H. V. (2000) The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis. *Nature* **408**, 613–615
78. Rochaix, J. D., Lemeille, S., Shapiguzov, A., Samol, I., Fucile, G., Willig, A., and Goldschmidt-Clermont, M. (2012) Protein kinases and phosphatases involved in the acclimation of the photosynthetic apparatus to a changing light environment. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 3466–3474
79. Peng, L., Fukao, Y., Fujiwara, M., Takami, T., and Shikanai, T. (2009) Efficient operation of NAD(P)H dehydrogenase requires supercomplex formation with photosystem I via minor LHCl in *Arabidopsis*. *Plant Cell* **21**, 3623–3640
80. Leister, D., and Shikanai, T. (2013) Complexities and protein complexes in the antimycin A-sensitive pathway of cyclic electron flow in plants. *Front. Plant Sci.* **4**, 161
81. Vallon, O., Bulte, L., Dainese, P., Olive, J., Bassi, R., and Wollman, F. A. (1991) Lateral redistribution of cytochrome b6/f complexes along thylakoid membranes upon state transitions. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8262–8266
82. Allred, D. R., and Staehelin, L. A. (1985) Lateral distribution of the cytochrome b(6)/f and coupling factor ATP synthetase complexes of chloroplast thylakoid membranes. *Plant Physiol.* **78**, 199–202
83. Anderson, J. M., and Andersson, B. (1982) The architecture of thylakoid membranes: lateral and transverse organization. *Trends Biochem. Sci.* **7**, 288–292
84. Kuras, R., and Wollman, F. A. (1994) The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga *Chlamydomonas reinhardtii*. *EMBO J.* **13**, 1019–1027
85. Stroebel, D., Choquet, Y., Popot, J. L., and Picot, D. (2003) An atypical haem in the cytochrome b(6)/f complex. *Nature* **426**, 413–418
86. Zhang, L., and Aro, E. M. (2002) Synthesis, membrane insertion and assembly of the chloroplast-encoded D1 protein into photosystem II. *FEBS Lett.* **512**, 13–18
87. Li, X. P., Björkman, O., Shih, C., Grossman, A. R., Rosenquist, M., Jansson, S., and Niyogi, K. M. (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**, 391–395
88. Kiss, A. Z., Ruban, A. V., and Horton, P. (2008) The PsbS protein controls the organization of the photosystem II antenna in higher plant thylakoid membranes. *J. Biol. Chem.* **283**, 3972–3978
89. Allorete, G., Tokutsu, R., Roach, T., Peers, G., Cardol, P., Girard-Bascou, J., Seigneurin-Berny, D., Petroustos, D., Kuntz, M., Breyton, C., Franck, F., Wollman, F. A., Niyogi, K. K., Krieger-Liszky, A., Minagawa, J., and Finazzi, G. (2013) A dual strategy to cope with high light in *Chlamydomonas reinhardtii*. *Plant Cell* **25**, 545–557
90. Boekema, E. J., van Roon, H., Calkoen, F., Bassi, R., and Dekker, J. P. (1999) Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Biochemistry* **38**, 2233–2239
91. Galka, P., Santabarbara, S., Khuong, T. T., Degand, H., Morsomme, P., Jennings, R. C., Boekema, E. J., and Caffarri, S. (2012) Functional analyses of the plant photosystem I-light-harvesting complex II supercomplex reveal that light-harvesting complex II loosely bound to photosystem II is a very efficient antenna for photosystem I in state II. *Plant Cell* **24**, 2963–2978
92. Zhang, H., Whitelegge, J. P., and Cramer, W. A. (2001) Ferredoxin:NADP+ oxidoreductase is a subunit of the chloroplast cytochrome b6/f complex. *J. Biol. Chem.* **276**, 38159–38165
93. Wagner, R., Aigner, H., and Funk, C. (2012) FtsH proteases located in the plant chloroplast. *Physiol. Plant* **145**, 203–214
94. Yoshioka, M., and Yamamoto, Y. (2011) Quality control of Photosystem II: where and how does the degradation of the D1 protein by FtsH proteases start under light stress?—Facts and hypotheses. *J. Photochem. Photobiol. B.* **104**, 229–235
95. Schuhmann, H., and Adamska, I. (2012) Deg proteases and their role in protein quality control and processing in different subcellular compartments of the plant cell. *Physiol. Plant* **145**, 224–234
96. Kato, Y., and Sakamoto, W. (2009) Protein quality control in chloroplasts: a current model of D1 protein degradation in the photosystem II repair cycle. *J. Biochem.* **146**, 463–469
97. Sun, X., Ouyang, M., Guo, J., Ma, J., Lu, C., Adam, Z., and Zhang, L. (2010) The thylakoid protease Deg1 is involved in photosystem-II assembly in *Arabidopsis thaliana*. *Plant J.* **62**, 240–249
98. Sun, X., Peng, L., Guo, J., Chi, W., Ma, J., Lu, C., and Zhang, L. (2007) Formation of DEG5 and DEG8 complexes and their involvement in the degradation of photodamaged photosystem II reaction center D1 protein in *Arabidopsis*. *Plant Cell* **19**, 1347–1361
99. Luciński, R., Misztal, L., Samardakiewicz, S., and Jackowski, G. (2011) The thylakoid protease Deg2 is involved in stress-related degradation of the photosystem II light-harvesting protein Lhcb6 in *Arabidopsis thaliana*. *New Phytol.* **192**, 74–86
100. 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
101. Mitra, S. K., Walters, B. T., Clouse, S. D., and Goshe, M. B. (2009) An efficient organic solvent based extraction method for the proteomic analysis of *Arabidopsis* plasma membranes. *J. Proteome Res.* **8**, 2752–2767
102. Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Bräutigam, A., Weber, A. P., and Izui, K. (2011) A plastidial sodium-dependent pyruvate transporter. *Nature* **476**, 472–475
103. Müller, M., Kunz, H., Schroeder, J. I., Kemp, G., Young, H. S., and Neuhaus, H. E. (2014) Decreased capacity of sodium export out of *Arabidopsis* chloroplasts impairs salt tolerance, photosynthesis and plant performance. *Plant J.* in press.
104. Pavón, L. R., Lundh, F., Lundin, B., Mishra, A., Persson, B. L., and Spetea, C. (2008) *Arabidopsis* ANTR1 is a thylakoid Na⁺-dependent phosphate transporter: functional characterization in *Escherichia coli*. *J. Biol. Chem.* **283**, 13520–13527
105. Thuswaldner, S., Lagerstedt, J. O., Rojas-Stütz, M., Bouhidel, K., Der, C., Leborgne-Castel, N., Mishra, A., Marty, F., Schoefs, B., Adamska, I., Persson, B. L., and Spetea, C. (2007) Identification, expression, and functional analyses of a thylakoid ATP/ADP carrier from *Arabidopsis*. *J. Biol. Chem.* **282**, 8848–8859
106. Gigolashvili, T., Geier, M., Ashykmina, N., Frerigmann, H., Wulfert, S., Krueger, S., Mugford, S. G., Kopriva, S., Haferkamp, I., and Flügge, U. I. (2012) The *Arabidopsis* thylakoid ADP/ATP carrier TAAC has an additional role in supplying plastidic phosphoadenosine 5'-phosphosulfate to the cytosol. *Plant Cell* **24**, 4187–4204

107. Ling, Q., Huang, W., Baldwin, A., and Jarvis, P. (2012) Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science* **338**, 655–659
108. Livingston, A. K., Cruz, J. A., Kohzuma, K., Dhingra, A., and Kramer, D. M. (2010) An *Arabidopsis* mutant with high cyclic electron flow around photosystem I (hcef) involving the NADPH dehydrogenase complex. *Plant Cell* **22**, 221–233
109. Page, M. L., Hamel, P. P., Gabilly, S. T., Zegzouti, H., Perea, J. V., Alonso, J. M., Ecker, J. R., Theg, S. M., Christensen, S.K., and Merchant, S. (2004) A homolog of prokaryotic thiol disulfide transporter CcdA is required for the assembly of the cytochrome b6f complex in *Arabidopsis* chloroplasts. *J. Biol. Chem.* **279**, 32474–32482
110. Choquet, Y., and Vallon, O. (2000) Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie (Paris)* **82**, 615–634