Proteome Analysis of the Rice Etioplast

METABOLIC AND REGULATORY NETWORKS AND NOVEL PROTEIN FUNCTIONS*⑤

Anne von Zychlinski‡§¶, Torsten Kleffmann‡§‖, Nandini Krishnamurthy,**
Kimmen Sjölander, Sacha Baginsky‡ ‡‡, and Wilhelm Gruissem‡**

We report an extensive proteome analysis of rice etio-plasts, which were highly purified from dark-grown leaves by a novel protocol using Nycodenz density gradient centrifugation. Comparative protein profiling of different cell compartments from leaf tissue demonstrated the purity of the etioplast preparation by the absence of diagnostic marker proteins of other cell compartments. Systematic analysis of the etioplast proteome identified 240 unique proteins that provide new insights into heterotrophic plant metabolism and control of gene expression. They include several new proteins that were not previously known to localize to plastids. The etioplast proteins were compared with proteomes from *Arabidopsis* chloroplasts and plastid from tobacco Bright Yellow 2 cells. Together with computational structure analyses of proteins without functional annotations, this comparative proteome analysis revealed novel etioplast-specific proteins. These include components of the plastid gene expression machinery such as two RNA helicases, an RNase II-like hydrolytic exonuclease, and a site 2 protease-like metalloprotease all of which were not known previously to localize to the plastid and are indicative for so far unknown regulatory mechanisms of plastid gene expression. All etioplast protein identifications and related data were integrated into a data base that is freely available upon request. *Molecular & Cellular Proteomics 4:1072–1084, 2005.*

Plastids are plant cell organelles that have essential biosynthetic and metabolic activities. These include photosynthetic carbon fixation and synthesis of amino acids, fatty acids, starch, and secondary metabolites such as pigments. Although plastids lost their autonomy and transferred most of their genes to the nucleus during evolution (1), they have retained a small genome encoding ~90 proteins. Different plastid types develop in a tissue-specific manner (for a detailed review on plastid biogenesis, see Ref. 2). According to their structure, pigment composition (color), and functional differentiation, plastids are classified as elaioplasts that are found in seed endosperm, chromoplasts in fruits and petals, amyloplasts in roots, etioplasts in dark-grown seedlings, and chloroplasts in photosynthetically active tissues (3). These specialized plastids types are typically the result of a differentiation program that is controlled by the cell and tissue type but also by environmental factors.

Perhaps the best understood example of plastid differentiation is the light-dependent conversion of etioplasts into chloroplasts. After exposure of dark-grown seedlings to light etioplasts differentiate into photosynthetically active chloroplasts within a few hours. Chloroplast differentiation is accompanied by the assembly of the thylakoid membrane-localized electron transport system, which requires proteins encoded by genes in both nuclear and chloroplast genomes (4, 5). Although chloroplast differentiation has been investigated in detail for many years, the molecular mechanisms that control the differentiation are not yet fully understood. Also information is still limited on the metabolic pathways that distinguish plastids in different heterotrophic tissues and from photosynthetically active chloroplasts.

Using proteome analysis we examined the global state of protein expression in etioplasts to establish comprehensive information on complex metabolic and regulatory networks that function in a heterotrophic plastid. Proteome analysis has become an indispensable source of information about protein expression, splice variants, and erroneous or incomplete prediction of gene structures in data bases. The analyses of cell organelle proteomes provide additional important information about protein localization and pathway compartmentalization. Most of the currently available plastid proteome information that provides new insights into organelle-specific metabolic functions has been reported from autotrophic chloroplasts (6–14). Here we report a novel protocol for the isolation of highly purified rice etioplasts and subsequent systematic analysis of the etioplast proteome. The proteins we identified reveal a complex etioplast-specific metabolism and novel regulatory functions in plastids during heterotrophic growth and early differentiation of the autotrophic chloroplast. Our results represent an important first step toward an integrated view of plastid metabolism and differentiation processes.

EXPERIMENTAL PROCEDURES

Plant Material—For each plastid isolation 200 g of rice (*Oryza sativa*, japonica cultivar group) seeds were washed in 5% sodium hydrochloride solution for 10 min, rinsed four times with water, and swollen overnight. The seeds were transferred to vermiculite watered

with half-concentrated Murashige and Skoog medium. Seedlings were grown in the dark at a constant 29 °C for 10 days.

Isolation and Purification of Etioplasts—Plastids were purified by a combination of two consecutive density gradient centrifugations using Nycodenz (Axis-Shield PoC AS, Oslo, Norway) as the density gradient medium and several differential centrifugation steps. Each step of the isolation procedure was carried out at 4 °C. Rice shoots were harvested and cut into 5–10-mm lengths. In batches of 50 g, the plant material was homogenized in 500 ml of etioplast isolation solution (E.I.S.)[1] containing 10 mM HEPES/KOH, pH 7.8, 2 mM EDTA, 2 mM MgCl₂, 1 mM tetrasodium pyrophosphate, 600 mM sorbitol, 0.2% (w/v) BSA (Fraction V, Sigma) with a Waring blender (three consecutive bursts at low speed and three at high speed). The homogenate was filtered through two layers of Miracloth (Calbiochem). The homogenization step was repeated once. The pooled filtrates were refiltered through four layers of Miracloth. The filtrate was subsequently centrifuged for 4 min at $200 \times g$ to remove cellular debris. The supernatant was recentrifuged for 10 min at $8000 \times g$. All centrifugation steps were carried out at 4 °C with a Sorvall RC5C centrifuge (DuPont). The pellets containing the plastids were carefully resuspended in E.I.S. and subjected to Nycodenz density gradient centrifugation. The supernatant of the differential centrifugation was centrifuged at $16,000 \times g$ for 20 min, and the resulting pellet containing a mixture of various cell organelles was put aside for further analysis.

A 50% (w/v) Nycodenz stock solution (containing 10 mM HEPES/KOH, pH 7.8, 2 mM EDTA, 2 mM MgCl₂, 1 mM tetrasodium pyrophosphate, 5 mM DTT) was diluted with E.I.S. (plus 5 mM DTT) to the required Nycodenz concentration and an osmolality of 500–550 mosM. The plastid suspension was adjusted with 50% Nycodenz stock solution to a final Nycodenz concentration of 30%. Five milliliters of the suspension were loaded in a 30-ml Corex tube (Number 8445). For 200 g of rice seeds we used four Corex tubes for the first Nycodenz gradient. The Nycodenz step gradient consisting of 6 ml of 25%, 8 ml of 20%, 6 ml of 15%, and 3 ml of 10% Nycodenz was then carefully casted on top of the organelle suspension. The gradient was centrifuged in a Sorvall HB-4 swinging bucket rotor at $8000 \times g$ for 45 min. Two yellowish bands (bands 2 and 3) at the interface of 20–15 and 25–20% Nycodenz and four pale grayish bands (bands 0, 1, 4, and 5) at the interface of 0–10, 10–15, and 25–30% and in the pellet of slightly higher than 30% Nycodenz were visible. All bands were examined by light microscopy using an Axioplan 2 microscope (Zeiss, Wetzikon, Switzerland). Bands 2 and 3 contained the highest amount of plastids and were used for further purifications. Bands 0, 1, 4, and 5 were diluted 3-fold with E.I.S. and centrifuged at $16,000 \times g$ for 20 min. The resulting pellets containing a mixture of various organelles were stored for additional analysis. Bands 2 and 3 were pooled and diluted 3-fold (v/v) with E.I.S. plus 5 mM DTT and subjected to differential centrifugation carried out in an SS34 rotor as described below.

1) The organelle suspension was centrifuged for 5 min at $8000 \times g$ to remove residual Nycodenz. The resulting pellet was resuspended in a maximum of 20 ml of E.I.S. per centrifuge tube and used for the second centrifugation step. 2) The second centrifugation was carried out for 10 min at $500 \times g$. The resulting pellet was resuspended as before and used for the third centrifugation step. 3) The third centrifugation was carried out for 15 min at $500 \times g$, resulting in the first plastid pellet.

To increase the yield the supernatant from the second centrifugation was recentrifuged for 10 min at $500 \times g$. The resulting pellet was resuspended, and the centrifugation step was repeated for 15 min at $500 \times g$, resulting in the second plastid pellet. Both pellets were combined and subjected to the second Nycodenz gradient centrifugation using the same composition as described for the first gradient (see above). Due to the low amount of resulting material we used only one centrifuge tube for the second gradient. After the second gradient bands 2 and 3 were collected separately, diluted 3-fold (v/v) with E.I.S., and centrifuged for 5 min at $8000 \times g$ to remove residual Nycodenz. The pellets were resuspended in a maximum of 20 ml of E.I.S. and centrifuged using an SS34 rotor for 10 min at $2000 \times g$. The final pellets were stored at −80 °C.

Solubilization of Proteins for SDS-Polyacrylamide Gel Electrophoresis—The pellets from the organelle fractionation, *i.e.* the pellet from band 2 of the second density gradient (plastid fraction) as well as the pellets from the supernatant of the first centrifugation step and from bands 0, 1, 4, and 5 after the first density gradient ("unclean" fractions) were resuspended in SDS sample buffer (0.675 mM Tris/HCl, pH 6.8, 10% glycerol, 20% 2-mercaptoethanol, bromphenol blue) and incubated at 30 °C for 30 min. Any insoluble material was pelleted for 10 min at $16,000 \times g$. Two hundred micrograms of solubilized proteins were directly subjected to SDS-PAGE by loading 50 μg/lane onto 10-cm-long homogeneous 12% polyacrylamide gels. After electrophoresis the gels were cut into 10 segments along the lane. Proteins in each gel segment were immediately subjected to in-gel tryptic digest as described previously (15) using sequencing grade modified trypsin (Promega) in a ratio of one part trypsin to 10 parts of protein. Proteins were digested overnight at 26 °C. After elution, tryptic peptides were lyophilized to dryness and stored at −80 °C until analysis.

Protein Identification by LC-MS/MS—Tryptic peptides of each fraction were resuspended in 5 μl of 5% ACN and 0.2% formic acid in water and loaded on laboratory-made silica capillary columns (inner diameter of 75 μm, length of 9 cm; BGB Analytik AG, Böckten, Switzerland) packed with C₁₈ reversed phase material (Magic C₁₈ resins; 5 μm, 200-Å pore; Michrom BioResources, Auburn, CA). The peptide mixture was separated and eluted by a gradient from 5 to 65% ACN over 2 h followed by an increase up to 80% during an additional 15 min. The flow rate at the tip of the column was adjusted to ~200 nl/min. LC was coupled on line to an LCQDeca XP ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a nanospray source. Mass analysis was performed with a spray voltage of 2.0–2.5 kV and one MS full scan followed by three data-dependent MS/MS scans of the three most intense parent ions. The dynamic exclusion function was enabled to allow two measurements of the same parent ion during 1 min followed by an exclusion duration of 1 min.

Analysis and Interpretation of Mass Spectrometric Data—MS/MS data were interpreted according to the standards put forward by Carr and colleagues (16). The SEQUEST software was used (Thermo Finnigan) to search the *O. sativa* protein data base (TIGR: The Institute of Genome Research, www.tigr.org/) including the plastid- and mitochondria-encoded proteins (download from May 5, 2004). dta files were created by the SEQUEST software for every MS/MS scan with a total ion count of at least 5×10^4, minimal peak count of 35, and a precursor ion mass in the range of 300–2000 *m/z*. Data were searched against the data base indexed for speed restricting the search to tryptic peptides without modifications (carboxyamidomethylated cysteines and oxidized methionines). To exclude any false

---

[1] The abbreviations used are: E.I.S., etioplast isolation solution; 2-D, two-dimensional; MIP, major intrinsic protein; BY-2, Bright Yellow 2; OPPP, oxidative pentose phosphate pathway; PPT, phosphoenolpyruvate/phosphate translocator; TPT, triose phosphate/phosphate translocator; PEP, phosphoenolpyruvate; POR (A and B), protochlorophyllide oxidoreductase (isoforms A and B); Tic, translocon of the inner envelope membrane of chloroplasts; Toc, translocon of the outer envelope membrane of chloroplasts; Tim, translocon of the inner membrane of mitochondria; CSP41, 41-kDa chloroplast stem-loop-binding protein; S2P, site 2 protease; SREBP, sterol regulatory element-binding protein; E-value, Expect value.

positive protein identification we manually interpreted each SEQUEST output by filtering the peptide hits using stringent hierarchical criteria as described previously (12). In brief, we accepted cross-correlation scores (Xcorr) of at least 2.5 for doubly and 3.5 for triply charged ions. Grouping of at least four significant peptide hits of the same protein was rated as significant protein identification. MS/MS spectra of the other protein hits were visually examined for a correct peak assignment considering criteria described previously (12). For peptide identifications with a ΔCN (normalized difference in correlation score, giving the differences between the front ranking and the following possible hit) lower than 0.1, the spectra of lower ranking hits were also examined. Identifications with a ΔCN of 0.0 resulting from different members of protein families, isoforms, or redundant data base entries that cannot be distinguished by the identified peptides are given in Supplemental Table 1B. In one case where we based a biological conclusion on a single hit protein identification, we searched the dta file with MASCOT (www.matrixscience.com) using the default parameters (error tolerance of 2 for the parent mass and 0.8 for the daughter ions) to confirm the SEQUEST identification. This MS/MS spectrum was identified with a MASCOT score of 78 as a peptide from the site 2 metalloprotease supporting the SEQUEST result (Supplemental Fig. 1).

Two-dimensional (2-D) PAGE—For 2-D PAGE plastid pellets of bands 2 and 3 were resuspended each in approximately 100 μl of solubilization buffer containing 40 mM Tris base, 7 M urea, 2 M thiourea, 2% CHAPS, 0.5% Brij 35, 0.4% carrier ampholytes, 2 mM tributyl phosphine (Fluka, Buchs, Switzerland), 20 mM DTT, complete EDTA-free protease inhibitor mixture (Roche Applied Science) resulting in a protein concentration of at least 1 μg/μl. Any insoluble material was pelleted for 30 min at 20,000 × g. For the first dimension 100 μg of protein were loaded onto 24-cm-long strips with an immobilized linear pH gradient from 4–7 (Bio-Rad) by in-gel rehydration. The rehydration was performed overnight in solubilization buffer without Tris base according to the manufacturer's instructions. Proteins were focused using the IPGphor (Amersham Biosciences) by the following voltage gradient: 2 h at 300 V, 1 h up to 600 V, 1 h up to 1000 V, 1 h up to 4000 V, and for a total of approximately 80,000 V-h at 4000 V. Focused strips were stored at −80 °C. The second dimension was performed in laboratory-made homogeneous 12% polyacrylamide gels using the Ettan Dalt 2 unit (Amersham Biosciences). Proteins were detected by SYPRO® Ruby staining (Molecular Probes Europe BV, Leiden, The Netherlands) and scanned with a Typhoon 9400 scanner (Amersham Biosciences). 2-D PAGE electropherograms were analyzed and compared with the ProteomWeaver software (Definiens, Munich, Germany).

Bioinformatic Analysis of Discovered Proteins—All relevant data of the 240 identified etioplast proteins were deposited in a relational data base designated as PLprot. This data base is accessible at www.pb.ethz.ch/proteomics and also contains information about the *Arabidopsis* chloroplast proteome as described previously (12). The TargetP program (www.cbs.dtu.dk/services/TargetP/) was used to predict subcellular protein targeting for plastids, mitochondria, the secretory pathway, and any other localization. For the identification of protein homologies the BLAST searches were conducted as described previously (17) against either the NCBI non-redundant data base (www.ncbi.nlm.nih.gov/BLAST/) or the Swiss-Prot data base using the BLAST program (bio.thep.lu.se/). Domain and structure homologies were analyzed either by the conserved domain search at NCBI (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using the Conserved Domain Database version 2.00 or with phylofacts (phylogenomics.berkeley.edu/resources/).

To provide more informative annotations of proteins without annotated function we further attempted informatic approaches as described previously (12). Briefly we performed a four-step procedure with the first two steps being essentially identical to those used in the

functional classification of the human genome (18). The third step focused on the prediction of structural domains in the proteins using fold recognition methods for the critical assessment of protein structure prediction (CASP) experiments (19). In the fourth step we finally used publicly available software tools (www.tigr.org/, www.ncbi.nlm.nih.gov/, www.cbs.dtu.dk/services/TargetP/, www.sanger.ac.uk/Software/Pfam/index.shtml, phylogenomics.berkeley.edu/resources, scop.berkeley.edu, www.rcsb.org/pdb, bioweb.pasteur.fr/seqanal/interfaces/toppred.html, bio.thep.lu.se/, and bibiserv.techfak.uni-bielefeld.de/dialign/) to complement our structure and function prediction analyses. A detailed description of the complete procedure was recently published for the annotation of protein functions for *Arabidopsis* chloroplast proteins (12).

RESULTS AND DISCUSSION

Isolation of Highly Purified Rice Etioplasts Requires Novel Density Gradient Centrifugation Methods—Large scale analysis of the etioplast proteome has been constrained by the lack of suitable strategies to isolate highly purified etioplasts from dark-grown plant tissue. This is of particular concern because protein detection has become more sensitive. Minor contaminations with proteins from other cell organelles might therefore result in misinterpretation of the proteome data.

To reduce ambiguities in the assignment of proteins to the etioplast proteome, we first tested different established plastid isolation methods for the purification of rice etioplasts. Sucrose density gradient centrifugation yielded plastids contaminated with other cell organelles and particles (data not shown) and exposed plastids to strong osmotic stress, which in general makes the isolation of intact plastids difficult. Percoll density gradient centrifugation resulted in the aggregation of rice etioplasts, which prevented their efficient separation from other cell organelles (data not shown). In contrast, a novel protocol that combines differential centrifugation and Nycodenz density gradient centrifugation (for details see "Experimental Procedures") allowed the isolation of intact and pure etioplasts from leaves of dark-grown rice seedlings.

After two consecutive Nycodenz density gradient centrifugations of leaf extracts two light yellow bands (bands 2 and 3) were visible at the interface of 15–20 and 20–25% Nycodenz. Four additional faint, light gray bands (bands 0, 1, 4, and 5) were visible at the interface of 0–10, 10–15, and 25–30% and in the pellet (Fig. 1A). Light microscopy analysis showed that bands 2 and 3 contained only etioplast-like organelles ranging in size between 2 and 4 μm (Fig. 1B). Differential interference contrast microscopy (Fig. 1C) revealed the etioplast-specific prolamellar body, suggesting that bands 2 and 3 contained highly purified etioplasts. Bands 0, 1, 4, and 5 (Fig. 1A) contained either etioplasts contaminated with other organelles or a mixture of different cell organelles and particles.

Protein Profiling Demonstrates High Purity of the Etioplast Preparation—Because light microscopy is not sufficient to exclude co-purification of other cell organelles during the Nycodenz density centrifugation, we analyzed the purity of etioplasts by comprehensive protein profiling using LC-MS/MS. We first identified diagnostic marker proteins from differ-
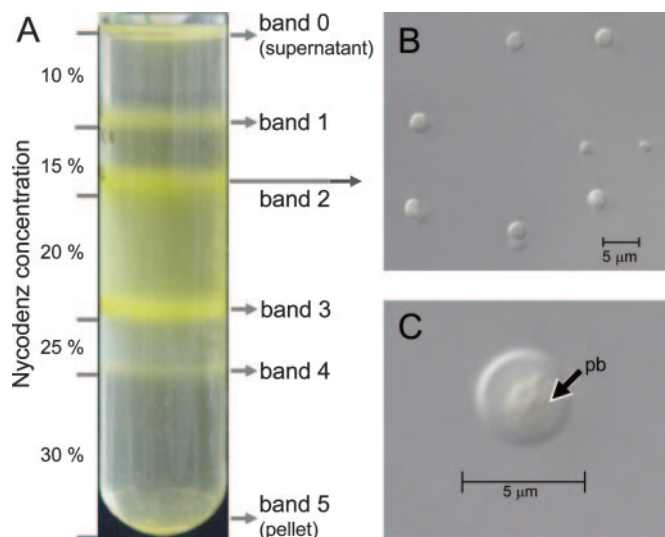
FIG. 1. **Isolation of rice etioplasts by Nycodenz density gradient centrifugation.** *A*, organelle fractions in the Nycodenz gradient. The first Nycodenz gradient centrifugation resulted in six bands, including the supernatant (band 0) and the pellet (band 5). Bands 2 and 3 were further purified as described under "Experimental Procedures." *B*, light microscopy of purified etioplasts. Following the second Nycodenz gradient band 2 contains a clean fraction of etioplasts ranging in size from 2 to 4 μm. *C*, differential interference contrast light microscopy reveals the etioplast-specific prolamellar body (*pb*).

ent cell organelles and the cytosol to monitor their presence in the different fractions obtained during the etioplast isolation procedure. Therefore we analyzed the supernatant of the first differential centrifugation and bands 0, 1, 4, and 5 (see Fig. 1*A*) from the first Nycodenz gradient by LC-MS/MS. Subsequently we analyzed the purified etioplasts isolated from band 2 of the second Nycodenz gradient and examined the protein complement of this band for the presence of marker proteins from other cell organelles. The purified etioplasts of bands 2 and 3 were later compared by differential display of their protein profiles using 2-D PAGE.

For a valid comparison of the proteins and their relative abundance in the different gradient fractions and the supernatant, we applied the same protein shotgun identification strategy to all fractions. Although determination of protein abundance in complex LC-MS/MS experiments is only semiquantitative, conclusions about relative protein abundance can be drawn from a pairwise comparison of the number of different peptides identified from the same protein (*i.e.* sequence coverage) in two different samples (for a review, see Ref. 20). Furthermore only those diagnostic marker proteins that are highly abundant in other cell compartments would be detected as contaminations in the etioplast fraction. Proteins that are enriched in the etioplast fraction compared with the other gradient fractions therefore can be considered true etioplast proteins or proteins that are specifically associated with the etioplast. This type of semiquantitative analysis is similar to a recently published protein profiling procedure using iso-

tope tagging of membrane proteins (LOPIT (21)). Both approaches allow a direct quantitative comparison of proteins from different cell organelles in the homologues system. This way, the enrichment of proteins by an organelle isolation procedure can be monitored. Together the described comparison of protein quantities in different protein fractions is more accurate to establish potential contaminations than cross-comparisons to different proteome studies that were performed with different biological material.

We identified 725 proteins from all gradient fractions and the supernatant. Two hundred and forty different proteins were identified from etioplasts in band 2 (Supplemental Table 1, A and B), and 579 were identified from all other fractions (94 identified proteins of all other fractions were overlapping with band 2). All identified proteins were examined for their subcellular localization. We accepted those proteins as true plastid proteins that (i) are encoded in the plastid genome or have a predicted plastid transit peptide (22), (ii) are annotated as plastid proteins in the TIGR data base, or (iii) are homologues of chloroplast proteins as annotated in the Swiss-Prot or NCBI data bases. Similarly we accepted proteins as targeted to other compartments according to the TIGR annotation or based on strong sequence homology to proteins with reported functions in mitochondria, peroxisomes, glyoxysomes, vacuoles, nuclei, the endoplasmic reticulum, the cytosol, or the extracellular space.

Of the set of 579 proteins from all other fractions except band 2 and band 3 etioplasts we considered 256 as potential diagnostic marker proteins because they were identified by at least three different peptides (Supplemental Table 2). Of these 256 proteins, 183 function in non-plastid compartments or other organelles or are of unknown subcellular localization (Supplemental Table 2). The distribution of these 183 diagnostic marker proteins among the different gradient fractions of the etioplast preparation is shown in Fig. 2, *A* and *B*. Using the criteria discussed above, we considered 216 of the 240 proteins identified from band 2 etioplasts as true plastid proteins. In addition, none of the abundant diagnostic marker proteins from other cell organelles was detected in the 240 identified etioplast proteins (Fig. 2*A*). This analysis clearly demonstrates that band 2 etioplasts were efficiently separated from all other cell organelles.

Assignment of Aconitase 1 and Proteins of Unknown Localization to Rice Etioplasts—Three proteins from the list of diagnostic marker proteins that were classified as "cytosolic" or "unknown subcellular localization" were overlapping with the 240 etioplast proteins (Fig. 2*B*). These proteins are annotated in the Swiss-Prot or NCBI data bases as cytosolic aconitase 1 (aconitate hydratase 1; 4203.m00200), putative major intrinsic protein (MIP) family protein (8292.m00189), and AtMSH3 full-length cDNA (8123.m00150). Although aconitase 1 was reported to have a cytosolic localization (23), its function is tightly coupled with the glutamine synthetase/glutamine-2-oxoglutarate amidotransferase (glutamine synthetase/gluta-
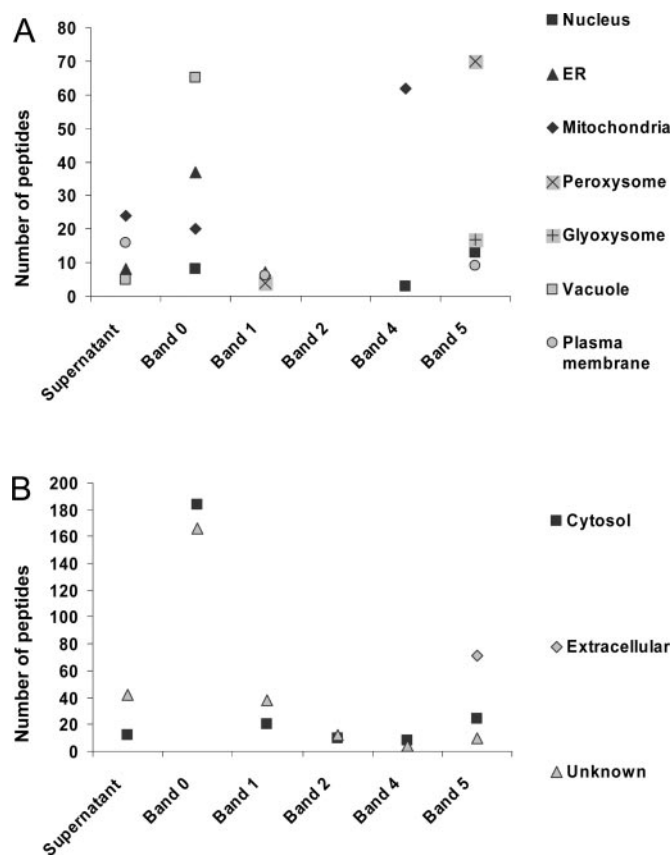
FIG. 2. **Distribution of diagnostic marker proteins in different gradient fractions during etioplast isolation.** *A*, diagnostic marker proteins from different cell organelles. *B*, diagnostic marker proteins from the cytosol, the extracellular space, and of unknown localization. The total number of peptides identified from all diagnostic marker proteins of each designated organelle is plotted against the fractions obtained during the etioplast isolation procedure (listed in Supplemental Table 2). None of the diagnostic marker proteins from other cell organelles was detected in the band 2 etioplast preparation. *ER*, endoplasmic reticulum.
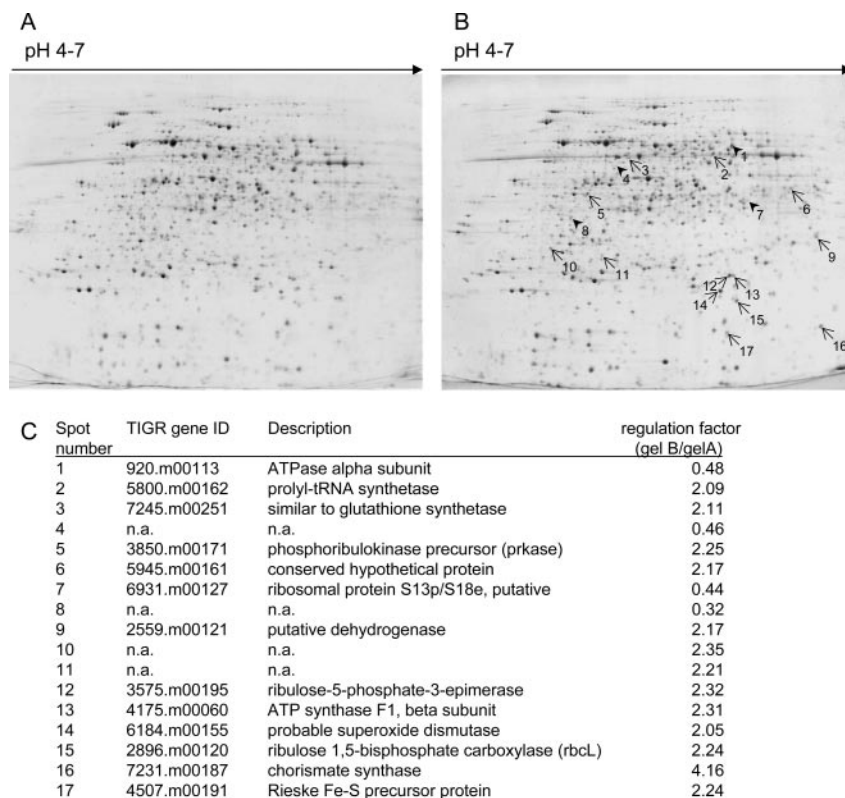
mate synthase) cycle activity (24). Because these essential steps of ammonia assimilation and amino acid biosynthesis are localized in the plastid, we propose that aconitase is functionally associated with the plastid. We consistently detected aconitase 1 with a high sequence coverage in band 2 of the etioplast preparation (10 unique peptides) compared with all other fractions of the Nycodenz gradient (with a maximum of eight unique peptides in band 4) (Supplemental Tables 1B and 2). This result suggests a specific enrichment of aconitase 1 by the etioplast isolation procedure. Additional evidence for a plastid association of aconitase was obtained from the proteome analysis of *Arabidopsis* chloroplast (12) and plastids of tobacco Bright Yellow 2 (BY-2) cells (25). In BY-2 plastids, aconitase was detected in the fraction containing peripheral membrane proteins that were obtained after washing the membranes with 8 M urea. Together these data suggest that aconitase is associated with the etioplast and provides precursors for the plastid amino acid metabolism.

For the other two proteins (MIP family protein 8292.m00189 and the unknown protein 8123.m00150, Fig. 2*B*) we found significant homologies to proteins that were previously identified in *Arabidopsis* chloroplasts (8292.m00189: At1g01620 and At3g61430 with E-values of 2e-106 and 7e-104; 8123.m00150: At1g01790 with an E-value of 4e-121) (10–12). Sequence and domain homologies of the MIP protein reveal potential functions in solute transport that would be consistent with its localization to the plastid envelope membrane system. Protein structure prediction using phylofacts (phylogenomics.berkeley.edu/resources/) revealed significant homologies to prokaryotic GlpF (glycerol uptake facilitator and related permeases; E-value of 1.65e-36). The bacterial protein is a transporter of small solutes and carbohydrates, suggesting that the plastid MIP protein has a similar function in the transport of carbohydrates between the plastid and the cytosol. Together the absence of abundant diagnostic marker proteins from other cell organelles and the cytosol suggests that the etioplasts we used for our subsequent proteome analysis were of high purity and did not contain other co-purifying cell organelles.

Several Etioplast Proteins Are Not Predicted by TargetP for Plastid Localization—Of the 240 proteins identified in band 2 etioplasts, 16 are plastid-encoded proteins (rice plastid, NCBI genome data base). TargetP predicts a plastid transit peptide for 168 (75%) of the 224 nuclear encoded proteins, whereas 24 proteins are predicted to localize to "any other location," 22 are predicted to localize to "mitochondria," and 10 are predicted to localize to the "secretory pathway" (Supplemental Table 1). An incorrect targeting prediction for plastid proteins can have several reasons, including the lack of a transit peptide in proteins targeted to the outer envelope membrane or the presence of non-canonical transit sequences (for reviews, see Refs. 26 and 27). Moreover dual targeting of proteins to plastids and other subcellular compartments has been reported for several proteins, further compounding the problem to predict plastid targeting (Refs. 28–30; and for a review, see Ref. 31). We identified 32 proteins that have no predictable plastid transit peptide but that were assigned as true plastid proteins by one of the above discussed criteria. Twenty-four of the etioplast proteins we identified do not have a predicted transit peptide and do not fit into one of the categories that classify plastid proteins (Supplemental Table 1A, compare columns "TargetP prediction" and "Verified subcellular localization"). This group of proteins is particularly important because it may represent proteins that potentially comprise novel plastid functions and/or protein import mechanisms.

Different Etioplast-like Organelles Can Be Separated by Nycodenz Density Gradient Centrifugation—After Nycodenz density gradient centrifugation we detected 17 plastid-assigned proteins exclusively in gradient fractions other then band 2 (see Supplemental Table 2, column "Notes"). This observation suggests that different etioplast-like organelles

FIG. 3. **Comparison of different etioplast isolates by protein profiles using 2-D PAGE.** *A*, etioplasts isolated from band 2 after the second Nycodenz density gradient centrifugation (Fig. 1). *B*, etioplasts isolated from band 3 after the second Nycodenz density gradient centrifugation (Fig. 1). The depicted 2-D PAGE electropherograms are average gel images created by the Proteom-Weaver software. The gel images display the average spot intensity from four (*A*) and two (*B*) individual gels. Using a regulation factor of 2 as threshold, only negligible differences were detected between band 2 and band 3 etioplasts (*arrows* point to protein spots up-regulated and *arrowheads* point to protein spots down-regulated in band 3 etioplasts). *C*, identification of up- and down-regulated proteins. The proteins most significantly up- and down-regulated (with a minimal spot intensity of at least 5% of the most intense spot) were identified by LC-ESI-MS/MS. *n.a.*, not analyzed.

Spot number	TIGR gene ID	Description	regulation factor (gel B/gelA)
1	920.m00113	ATPase alpha subunit	0.48
2	5800.m00162	prolyl-tRNA synthetase	2.09
3	7245.m00251	similar to glutathione synthetase	2.11
4	n.a.	n.a.	0.46
5	3850.m00171	phosphoribulokinase precursor (prkase)	2.25
6	5945.m00161	conserved hypothetical protein	2.17
7	6931.m00127	ribosomal protein S13p/S18e, putative	0.44
8	n.a.	n.a.	0.32
9	2559.m00121	putative dehydrogenase	2.17
10	n.a.	n.a.	2.35
11	n.a.	n.a.	2.21
12	3575.m00195	ribulose-5-phosphate-3-epimerase	2.32
13	4175.m00060	ATP synthase F1, beta subunit	2.31
14	6184.m00155	probable superoxide dismutase	2.05
15	2896.m00120	ribulose 1,5-bisphosphate carboxylase (rbcL)	2.24
16	7231.m00187	chorismate synthase	4.16
17	4507.m00191	Rieske Fe-S precursor protein	2.24

exist in etiolated rice leaves. One possible explanation is a reported developmental gradient of plastids along the leaf axis of monocotyledonous plants (32). To better understand the relationship between band 2 etioplasts and other etioplast-like organelles, we compared the protein profile of band 2 etioplasts with that of band 3 etioplasts. Band 3 contains highly purified organelles that are identical in size and shape to band 2 etioplasts as judged by light microscopy (see Fig. 1). Band 2 and band 3 etioplasts were analyzed by 2-D PAGE using at least two replicates of each protein sample. All replicate gels were integrated into one average gel using the ProteomWeaver software (for details see "Experimental Procedures"). On the basis of a threshold factor of 2 for different protein staining intensity, the protein profiles of band 2 and band 3 etioplasts are virtually identical (Fig. 3, *A* and *B*). Some protein spots differed in their staining intensity, but most of these differences derive from protein fragments of proteins such as ribulose-1,5-bisphosphate carboxylase large subunit and ATPase β subunit (the molecular masses of the fragments determined by 2-D gel analysis were significantly lower than the predicted masses of the full proteins) (Fig. 3*C*). Our analysis confirms that both band 2 and band 3 contain highly purified etioplasts. Further detailed analysis of *e.g.* lipid or carbohydrate composition will be necessary to explain the different densities of rice etioplasts.

Identified Etioplast Proteins Belong to Metabolic Modules Characteristic of Heterotrophic Leaves—Although shotgun proteome analysis does not provide information about en-zyme activity, the protein profile provides a comprehensive view of the metabolic pathways in the etioplast. We identified an enzyme complement that is consistent with an etioplast-specific metabolic network in a heterotrophic plant tissue (Fig. 4). Although photosynthetically active (autotrophic) chloroplasts produce reductants, energy-rich cofactors, and metabolite precursors for anabolic metabolism and CO_2 fixation, non-green (heterotrophic) plastids either import these molecules or produce them via oxidative metabolism in the organelle. Glucose 6-phosphate, for example, is imported into non-photosynthetic plastids and utilized by the oxidative pentose phosphate pathway (OPPP) to generate reducing power (NADPH) and metabolite precursors for essential anabolic pathways (for reviews, see Refs. 3 and 33). Another source of reducing equivalents in non-photosynthetic plastids is glycolysis, which utilizes glucose 6-phosphate or a partial glycolysis module that only utilizes phosphoenolpyruvate (PEP) to generate ATP and pyruvate. Pyruvate is further converted by the pyruvate dehydrogenase complex to acetyl-CoA and a reducing equivalent.

Although different non-photosynthetic plastids such as amyloplasts import cytosolic hexose phosphates for energy metabolism and starch synthesis, other reports suggest that barley etioplasts import cytosolic triose phosphates that can be converted into hexose phosphates in the plastid (Ref. 34; summarized in Refs. 3 and 35). The import of triose phosphates in the dark requires the activity of fructose-1,6-bisphosphatase, which is activated by thioredoxin in the light
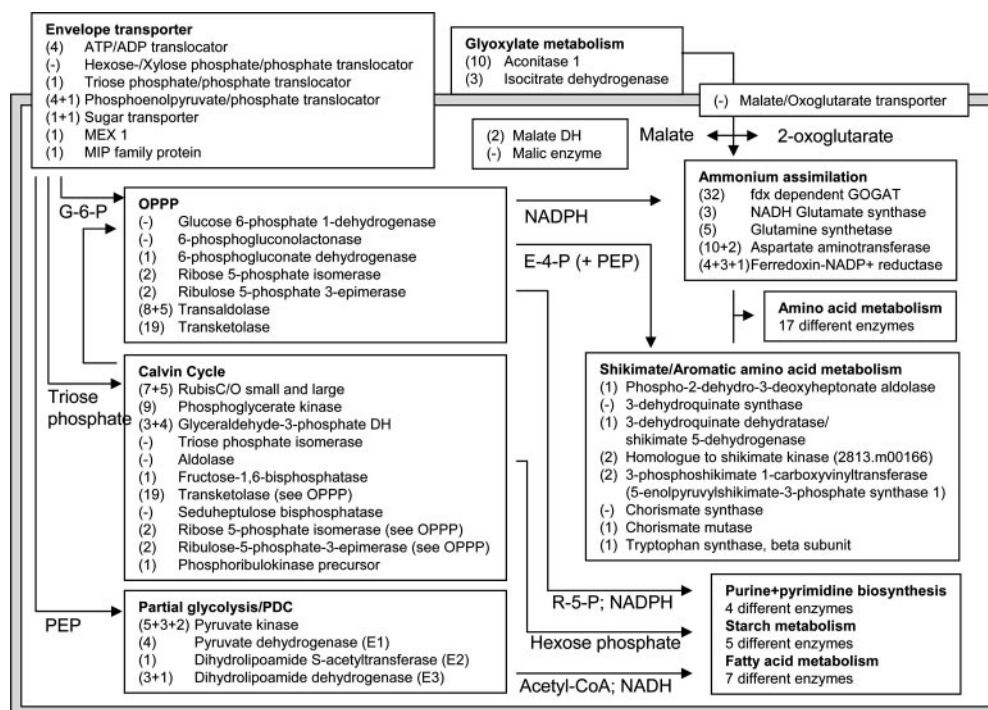
Envelope transporter
(4) ATP/ADP translocator
(-) Hexose-/Xylose phosphate/phosphate translocator
(1) Triose phosphate/phosphate translocator
(4+1) Phosphoenolpyruvate/phosphate translocator
(1+1) Sugar transporter
(1) MEX 1
(1) MIP family protein

Glyoxylate metabolism
(10) Aconitase 1
(3) Isocitrate dehydrogenase

(-) Malate/Oxoglutarate transporter

(2) Malate DH
(-) Malic enzyme

Malate ⇌ 2-oxoglutarate

G-6-P

OPPP
(-) Glucose 6-phosphate 1-dehydrogenase
(-) 6-phosphogluconolactonase
(1) 6-phosphogluconate dehydrogenase
(2) Ribose 5-phosphate isomerase
(2) Ribulose 5-phosphate 3-epimerase
(8+5) Transaldolase
(19) Transketolase

NADPH

E-4-P (+ PEP)

Ammonium assimilation
(32) fdx dependent GOGAT
(3) NADH Glutamate synthase
(5) Glutamine synthetase
(10+2) Aspartate aminotransferase
(4+3+1) Ferredoxin-NADP+ reductase

Amino acid metabolism
17 different enzymes

Triose phosphate

Calvin Cycle
(7+5) RubisC/O small and large
(9) Phosphoglycerate kinase
(3+4) Glyceraldehyde-3-phosphate DH
(-) Triose phosphate isomerase
(-) Aldolase
(1) Fructose-1,6-bisphosphatase
(19) Transketolase (see OPPP)
(-) Seduheptulose bisphosphatase
(2) Ribose 5-phosphate isomerase (see OPPP)
(2) Ribulose-5-phosphate-3-epimerase (see OPPP)
(1) Phosphoribulokinase precursor

Shikimate/Aromatic amino acid metabolism
(1) Phospho-2-dehydro-3-deoxyheptonate aldolase
(-) 3-dehydroquinate synthase
(1) 3-dehydroquinate dehydratase/
    shikimate 5-dehydrogenase
(2) Homologue to shikimate kinase (2813.m00166)
(2) 3-phosphoshikimate 1-carboxyvinyltransferase
    (5-enolpyruvylshikimate-3-phosphate synthase 1)
(-) Chorismate synthase
(1) Chorismate mutase
(1) Tryptophan synthase, beta subunit

PEP

Partial glycolysis/PDC
(5+3+2) Pyruvate kinase
(4) Pyruvate dehydrogenase (E1)
(1) Dihydrolipoamide S-acetyltransferase (E2)
(3+1) Dihydrolipoamide dehydrogenase (E3)

R-5-P; NADPH

Hexose phosphate

Acetyl-CoA; NADH

Purine+pyrimidine biosynthesis
4 different enzymes
Starch metabolism
5 different enzymes
Fatty acid metabolism
7 different enzymes

FIG. 4. **Schematic illustration of metabolic pathways in heterotrophic plastids.** The *number* of identified peptides for each identified protein is given in *parentheses*. Several numbers are listed if different protein isoforms were identified. Missing identifications of proteins depicted in the pathways are indicated by *dashes*. *E-4-P*, erythrose 4-phosphate; *G-6-P*, glucose 6-phosphate; *PDC*, pyruvate dehydrogenase complex; *R-5-P*, ribulose 5-phosphate; *GOGAT*, glutamine 2-oxoglutarate amidotransferase; *DH*, dehydrogenase.

(3). Although the proposed import of triose phosphates into etioplasts contradicts our current understanding of light-dependent regulation of plastid enzymes via reduced thioredoxin (for a review, see Ref. 36), our proteomic data support the view that etioplasts import triose phosphates rather than hexose phosphates. We identified two isoforms of a phosphoenolpyruvate/phosphate translocator (PPT) and a triose phosphate/phosphate translocator (TPT) in the etioplast but no xylulose 5-phosphate/phosphate translocator or glucose 6-phosphate/phosphate translocator (summarized in Fig. 4 and Supplemental Table 1, column "Functional category"). We also did not detect key enzymes of the glycolytic pathway, such as phosphoglyceromutase and enolase, which catalyze the two last steps in the conversion of hexose phosphates to PEP. On the other hand, the presence of two pyruvate kinase isoforms and all three subunits of the pyruvate dehydrogenase complex suggests that the combined activity of a partial glycolysis module and the pyruvate dehydrogenase complex is of specific importance for etioplast metabolism.

Another important source of reducing equivalents in non-photosynthetic plastids is the OPPP that utilizes glucose 6-phosphate (3). The oxidative branch comprises three enzymatic reactions that convert glucose 6-phosphate into ribulose 5-phosphate and generates two molecules of NADPH. The non-oxidative branch comprises the concerted action of ribose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase, and transaldolase. We identified only

6-phosphogluconate dehydrogenase from the oxidative branch, whereas all four enzymes of the non-oxidative branch of OPPP were identified. This is notable because the enzymes of the non-oxidative branch (with the exception of the transaldolase) are amphibolic enzymes of the OPPP and the Calvin cycle, both of which contribute carbon skeletons for anabolic metabolism, e.g. erythrose 4-phosphate as a precursor for chorismate biosynthesis. Thus, our proteome analysis is consistent with a view that the concerted activity of TPT and PPT and enzymes of the Calvin cycle together with the partial modules of glycolysis and OPPP allows the etioplast to effectively generate reducing power, ATP, and carbon skeleton precursors for its anabolic metabolism.

A comparison of the etioplast proteome reported here and proteins identified from non-photosynthetic BY-2 plastids (25) also suggests an etioplast-specific heterotrophic metabolism in dark-grown leaves. In BY-2 plastids no PPT or TPT was detected, whereas glucose 6-phosphate/phosphate translocator is highly abundant (25). Thus, BY-2 plastids appear to import preferentially hexose phosphates. Important Calvin cycle enzymes that were identified in etioplasts are absent or of low abundance in BY-2 plastids, confirming important metabolic differences between these two non-photosynthetic plastid types in heterotrophic tissues. These differences could be a consequence of the differentiation status of both plastid types. Although etioplasts differentiate into chloroplasts after illumination, BY-2 plastids remain non-photosynthetic in the

light and can only differentiate into amyloplast-like organelles after hormone depletion (37). It is also possible that the BY-2 plastid metabolism reflects the long term adaptation of the BY-2 cell culture to the carbohydrate supply from the growth medium. Glucose feeding can induce transport of glucose 6-phosphate into plastids (38), whereas the expression of TPT in tobacco seedlings decreased after external sucrose supply (39). Our results clearly show that the etioplasts have a specific metabolism that differs from other heterotrophic plastid types.

The Etioplast Proteome Analysis Reveals a Complex Anabolic Metabolism—The etioplast proteome reported here covers a wide variety of enzymes for essential and predominantly plastid-localized anabolic pathways that provide the plant cell with metabolites in the dark. Anabolic pathways in the etioplast require energy and reducing equivalents that are generated by the heterotrophic metabolism in the plastid and by import of ATP from the cytosol. In the rice etioplast we identified an ADP/ATP translocator, a MIP protein (see above), and several sugar transporters, including a homologue to *Arabidopsis* root cap1 protein that was recently reported as a maltose transporter, MEX1 (40). These transporters may play an important role in the shuttling of energy-rich molecules and metabolic intermediates between the cytosol and the etioplast.

Ammonia assimilation and synthesis of amino acids are one of the major biosynthetic functions of plastids. All enzymes of the glutamine synthetase/glutamine-2-oxoglutarate amidotransferase (glutamine synthetase/glutamate synthase) cycle were detected in the rice etioplast, including ferredoxin and ferredoxin-NADP(H) reductases. Furthermore several enzymes for the synthesis of threonine, lysine, methionine, histidine, the branched-chain amino acids, and nearly all enzymes of the shikimic acid pathway were detected (Fig. 4). Because shotgun proteomics preferentially identifies abundant proteins (41), we found only partial coverage of enzymes from other plastid-specific pathways. We identified several enzymes downstream of chorismate mutase or active in starch and fatty acid metabolism, however, confirming that other important pathways are also present and most likely active in the etioplast.

The tetrapyrrol pathway, for which we identified several enzymes in our analysis, produces protochlorophyllide as the end product of this pathway in the dark. Protochlorophyllide in etioplasts is stored in a ternary complex, the prolamellar body, together with protochlorophyllide oxidoreductase (POR) and NADPH (42, 43). We identified the two isoforms, POR A and POR B, in the rice etioplast together with two subunits of the magnesium chelatase (subunits chlI and chlH) but not ferrochelatase, which channels tetrapyrrols to the heme branch of the pathway. The results confirm that tetrapyrrol biosynthesis in etioplasts is channeled to protochlorophyllide and not to heme. This pathway is important for the photosynthetically inactive etioplast because intermediate products, *e.g.* Mg-protoporphyrin IX and Mg-protoporphyrin 6-monomethyl es-

ter play a fundamental role in signaling between plastid and nucleus by inhibiting the expression of nuclear genes for photosynthetic proteins (Ref. 44; for reviews, see Refs. 45–47).

The Etioplast Protein Import Machinery Reveals a Previously Unrecognized Complexity—Plastids import nuclear encoded proteins from the cytosol to support their essential metabolic activities. Most proteins of the translocon complex in the outer (Toc) and inner (Tic) envelope membrane are now well established (for reviews, see Refs. 48 and 49). We identified the transmembrane proteins Toc159, Tic110, and Tic40 and the soluble component pitrilysin (stromal processing peptidase) in rice etioplasts (Supplemental Table 1). Tic40 is noteworthy because this protein was not detected in a comprehensive proteome analysis of chloroplasts from 7-week-old *Arabidopsis* plants (12) but was identified in chloroplasts from younger plants (10, 11). Recent studies have shown that Tic40 is not essential for protein translocation (50). The described temporal expression of Tic40 and its detection in the etioplast suggest, however, that the protein may be necessary to modulate the transport mechanism in response to the changing demand for protein import.

In view of "non-canonical" targeting sequences in proteins imported into plastids (27) and the growing number of proteins with "dual targeting" to different organelles (Refs. 28–30; for a review, see Ref. 31), it is also possible that currently unknown proteins participate in plastid protein import. We identified one protein (3849.m00140) with similarity to the *Arabidopsis* Tim17/Tim22/Tim23 (translocon of the inner membrane of mitochondria) family of proteins (At5g24650 and At3g49560). Although the domain homology of 3849.m00140 to Tim17/Tim22/Tim23 proteins is poor (E-value of 0.00053 and 34.8% alignment coverage), its overall sequence similarity to the *Arabidopsis* proteins is significant (E-values of 2e-37 and 4e-33). The *Arabidopsis* protein At5g24650 was also identified in the envelope membrane of *Arabidopsis* chloroplasts (10, 11), supporting the localization of 3849.m00140 in the rice etioplast. The identification of a Tim17/Tim22/Tim23 family protein in the etioplast envelope membrane together with the identification of Tic40 in different plastids suggests a previously unrecognized complexity and dynamic composition of the plastid protein import machinery.

A Comparison of Proteomes from Different Plastid Types Reveals Novel Etioplast-specific Proteins—We searched the 224 identified nuclear encoded etioplast proteins for homologues in the *Arabidopsis* genome to identify proteins that are specific for rice. Only four proteins have no homologue in *Arabidopsis* if a BLAST E-value of 1e-20 is used as the cutoff. This group comprises a putative cystathionine β synthase domain protein (2451.m00094: E-value of 2e-5), a putative ribosomal protein S21 (2460.m00128: E-value of 6e-12), and two hypothetical proteins (3045.m00146: E-value of 2e-09; 3029.m00146: E-value of 3e-18). This result suggests that at the current level of coverage both *Arabidopsis* and rice plastid
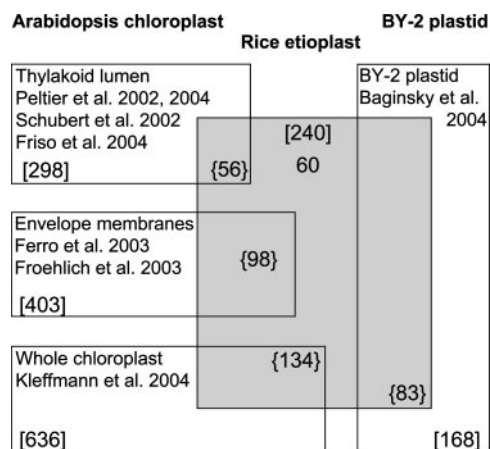
FIG. 5. **Identification of proteins that are prevalent in etioplasts.** Nuclear encoded etioplast proteins were BLAST-searched for homologue proteins that were previously identified in proteome studies of the *Arabidopsis* chloroplast and tobacco BY-2 plastids. Proteins with a BLAST E-value of *e*-20 or lower were considered as significant homologues. The *number* in *brackets* represents the total number of proteins identified in each study, and those in *braces* represent the overlap to the identified etioplast proteome. Sixty proteins were exclusively identified in the etioplast and therefore are most likely etioplast-specific or have a significant function for this heterotrophic plastid. *Peltier et al. 2002, 2004*, Refs. 8 and 14, respectively; *Schubert et al. 2002*, Ref. 9; *Friso et al. 2004*, Ref. 51; *Ferro et al. 2003*, Ref. 10; *Froehlich et al. 2003*, Ref. 11; *Kleffmann et al. 2004*, Ref. 12; *Baginsky et al. 2004*, Ref. 25.

proteomes are well conserved.

Different plastid types perform specialized functions in plant cells. It is therefore reasonable to assume that each plastid type comprises a proteome that reflects its function in a specific cellular context. To identify proteins that are prevalent in etioplasts compared with other plastid types, we compared the rice etioplast proteome to 1023 experimentally identified *Arabidopsis* chloroplast proteins from the complete chloroplast (12), the chloroplast envelope membrane (10, 11), and the thylakoid membrane and lumen (8, 9, 14, 51). We also included 168 proteins experimentally identified from tobacco BY-2 plastids (25). One hundred and eighty of the rice etioplast proteins have a significant homologue (E-value 1*e*-20 cutoff) reported in one of the above proteome studies (Fig. 5 and Supplemental Table 4). Most of these homologues (134) were identified in the complete *Arabidopsis* chloroplast (12). Although the etioplast does not contain a fully developed thylakoid membrane system, 56 etioplast proteins have significant homology to proteins found at the thylakoid membrane or in the thylakoid lumen (8, 9, 14, 51). These include characteristic luminal proteins such as oxygen-evolving complex proteins (two members of the PsbP family and PsbQ) and plastocyanin. In addition, several of these homologues function in redox regulation (*e.g.* thioredoxins and peroxiredoxins). The detection of proteins involved in photosynthesis in the non-photosynthetic etioplast is in line with early reports about the prethylakoid membrane system and especially photosystem II. Many of the protein subunits of photosystem II including PsbH, the oxygen-evolving complex proteins, and D2 are synthesized independently of light and are known to accumulate in etioplasts to basal levels (for a review, see Ref. 52).

Ninety-eight etioplast proteins are homologous to proteins reported from proteome studies of the chloroplast envelope membrane (10, 11). This suggests that many of the identified etioplast proteins potentially localize to the etioplast envelope membrane.

Together 71 of the identified etioplast proteins have no significant homologue to experimentally verified *Arabidopsis* chloroplast proteins. It is possible that these proteins are specific to non-photosynthetic (heterotrophic) plastids. We therefore searched for homologues in a set of 168 proteins identified from BY-2 plastids (Fig. 5). Only 11 of the 71 etioplast proteins are shared between the two heterotrophic plastid types from etiolated rice leaves and BY-2 cells. With the exception of magnesium chelatase subunit chlI (1485.m00199), the other shared proteins belong to amino acid synthesis pathways. At the current coverage of experimentally verified plastid proteomes, 60 (25%) of the identified rice proteins appear to have a higher relative abundance in etioplasts compared with other plastid types. Most of these 60 proteins are involved in gene expression and carbohydrate and amino acid metabolism. Twenty of these proteins currently have no predicted function and are therefore promising candidates for the discovery of new etioplast-specific activities in the plant cell (see Supplemental Table 4).

Structure Prediction Analysis of Etioplast-specific Proteins Suggests New Plastid Protein Functions and Improves Functional Annotation—Forty-six of all identified etioplast proteins currently have no annotation in the TIGR data base or reported function in plastids. To provide more informative functional annotations we analyzed these proteins using computational structure prediction methods and phylogenetic analyses (for details, see "Experimental Procedures"). Thirty-two of the 46 proteins contained high confidence matches to a functionally and/or structurally characterized domain (Supplemental Table 5). Phylogenomic analysis was applied to those sequences for which a sufficient number (20–30) of credible homologues could be identified. Analysis of the phylogenetic distribution of those 32 proteins shows that roughly 25% of them have homologues spanning all branches of the tree of life, 14% are restricted to eukaryotes, and 11% are restricted to plants. Not surprisingly, nearly 50% of all analyzed proteins have close homologues in cyanobacteria.

Several proteins were predicted to control gene expression and were analyzed in more detail. Two identified etioplast proteins (2927.m00126, RNB-like protein; and 5895.m00118, mRNA-binding protein-related) have restricted homology to proteins involved in RNA binding and processing. The RNB-like protein has a C-terminal sequence that is similar to the domain architecture of exoribonuclease II (E-value of 3*e*-50) and an overall sequence homology to a ribonuclease II family
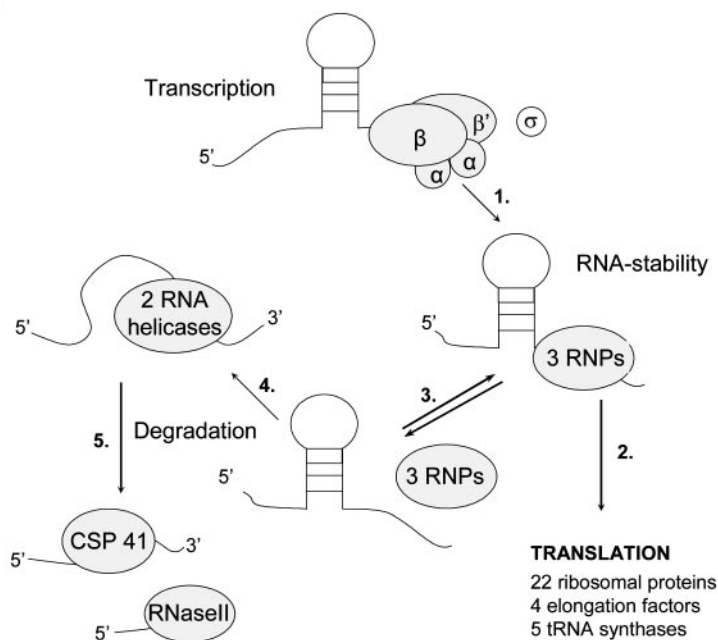
FIG. 6. **The etioplast proteome reveals a dark-specific RNA degradation pathway.** *A*, multiple sequence alignments of CSP41 from different species and the consensus sequence of NAD-dependent epimerase/dehydratase family proteins. Relevant motifs in the sequences of the identified etioplast RNA-binding protein (5895.m00118) were aligned with CSP41 of *Arabidopsis* (*ARATH*), spinach (*SPIOL*), tobacco (*TABAC*), *Synechocystis* (*SYNY*), and a consensus sequence for NAD-dependent epimerase/dehydratase family proteins (*consensus_ED*). *Arrowheads* point to three conserved glycines of NAD-dependent epimerase/dehydratase family proteins, and *arrows* point to amino acid residues conserved in CSP41 (for details see text). *B*, dark-specific RNA degradation pathway. The set of identified proteins (*gray circles*) allows us to reconstruct a complete dark-specific RNA degradation pathway in the etioplast (for a review, see Ref. 50). Following transcription by the PEP polymerase (*1.*), the mRNA can be stabilized for translation by RNA-binding proteins (*2.*) or directed to a degradation pathway (*3.*). The 3′ stem-loop structure of the mRNA is unwound by RNA helicases (*4.*) followed by endo- and exonucleolytic degradation (*5.*) of the cleavage products. α, β, and β', subunits of plastid-encoded RNA polymerase (PEP); RNPs, RNA-binding proteins.

protein from *Synechocystis* (E-value of *e*-103 using BLAST search against Swiss-Prot). This suggests that the RNB-like protein could function in the etioplast as a 3′–5′ exonuclease in plastid RNA degradation.

The mRNA-binding protein 5895.m00118 shows significant homologies to the NAD-dependent epimerase/dehydratase family of proteins (E-value of 2*e*-08, conserved domain search at NCBI). Certain NAD-dependent epimerase/dehydratases have reported RNA binding activities. Among them is 41-kDa chloroplast stem-loop-binding protein (CSP41), a plastid divalent metal-dependent endoribonuclease (Ref. 53; summarized in Ref. 54). In comparison to nucleotide-sugar epime-

rases, CSP41 contains only two of three conserved glycines in the nucleotide binding motif (53). The acidic residues Asn[71], Asp[89], and Asp[103] are important in divalent metal binding activity (Ref. 55; summarized in Ref. 54), and Ser[118], His[130], and His[142] are important for RNA catalysis (53). Multiple sequence alignments of the etioplast protein 5895.m00118 with CSP41 of *Arabidopsis* (T48103), spinach (AAC49424), tobacco (AAP87140), *Synechocystis* (BAA17464), and a consensus sequence for NAD-dependent epimerase/dehydratase family proteins (pfam01370.11, epimerase; NCBI conserved domain search) show significant sequence identities to all aligned CSP41 sequences (Fig. 6*A*). All conserved

amino acid residues of CSP41 listed above can be aligned but are partially absent from the consensus sequence of NAD-dependent epimerase/dehydratase family proteins. Together the strong overall sequence homology (E-value of 1e-133) and the significant motif similarities to spinach CSP41 suggest that the identified etioplast protein is a rice orthologue of spinach CSP41.

Based on the proteins we identified in etioplasts and that have been reported in the literature, we can propose a plastid RNA degradation pathway (Fig. 6*B*). The results support a model in which CSP41 is the non-site-specific key endonuclease that controls RNA half-lives in response to cellular and environmental conditions (56, 57). Interestingly we did not identify the known 3′ to 5′ exonuclease polynucleotide phosphorylase (58, 59) but an RNase II-like enzyme. RNase II is a hydrolytic exonuclease, whereas polynucleotide phosphorylase catalyzes phosphorolytic RNA degradation (58, 59). We suggest that hydrolytic RNA degradation is prevalent in etioplasts and is replaced by phosphorolytic RNA degradation in photosynthetically active chloroplasts. These fundamentally different plastid RNA degradation mechanisms would be consistent with the different energy metabolism and requirement of inorganic phosphate in heterotrophic and autotrophic plant tissues. In addition to endo- and exonucleases, we discovered two novel plastid RNA helicases. Their presence in etioplasts further supports the importance of RNA secondary structures in the regulation of RNA stability (60). RNA helicases unwind secondary structures to make them accessible to endonucleolytic or exonucleolytic cleavages. It is likely that CSP41 together with RNA helicases initiates the RNA degradation pathway in etioplasts, releasing short single-stranded RNA molecules that are then completely degraded by the RNase II like 3′–5′ exonuclease (Fig. 6*B*).

We also discovered a sterol regulator element-binding site 2 protease (SREBP, 2714.m00157) homologue that could regulate gene expression at the level of transcription (Supplemental Fig. 1). Site 2 proteases (S2Ps) belong to a group of transmembrane metallopeptidases that regulate gene expression in both prokaryotes and eukaryotes (61). Members of this protein family contain several transmembrane-spanning domains and the metallopeptidase-specific conserved metal binding motif HE$X$$X$H as well as a conserved C-terminal motif NX_nPX_nDG that is characteristic for S2Ps (61). Both motifs together form the active catalytic site at a junction adjacent to/or within the membrane (for reviews, see Refs. 62 and 63). Consistent with the characteristic structure of S2Ps, the etioplast protein 2714.m00157 has six predicted transmembrane domains (TopPred software at bioweb.pasteur.fr/seqanal/interfaces/toppred.html using the Kyte and Doolittle hydrophobicity scale and a core window size of 20) (Fig. 7). The HE$X$$X$H motif (HEIAH at amino acid position His[314]–His[318]) is localized in the second transmembrane domain, and the NX_nPX_nDG motif (NSIPAGELDG at position Asn[442]–Gly[451]) is at the end of the fourth transmembrane domain (Fig. 7). This topology en-
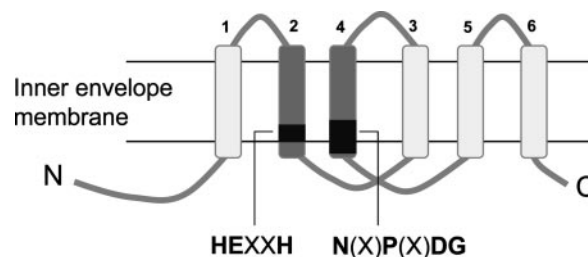

FIG. 7. **Suggested membrane topology of an etioplast S2P-like metallopeptidase.** On the basis of the Kyte and Doolittle hydrophobicity scale and a core window size of 20 the TopPred software (bioweb.pasteur.fr/seqanal/interfaces/toppred.html) predicted six transmembrane domains with the following topology: 1, Thr[256]–Asp[285]; 2, Leu[296]–Thr[325]; 3, Asp[358]–Ser[387]; 4, Gly[420]–Leu[449]; 5, Ile[467]–Gln[496]; 6, Asn[512]–Leu[541]. The two conserved domains HE$X$$X$H (His[314]–His[318]) and NX_nPX_nDG (Asn[442]–Gly[451]) are located outside or slightly inside of the second and fourth transmembrane domains as indicated. The membrane topology and location of the catalytic site suggested for the rice protein is similar to that described for the sporulation protein (SpoIVFB) of *B. subtilis* (61).

ables the formation of the catalytic site at or within the membrane, similar to the model for the sporulation protein SpoIVFB of *Bacillus subtilis* (61). SpoIVFB is a well characterized S2P-like membrane-embedded metalloprotease that activates the σ^K transcription factor by proteolytic cleavage of a membrane-associated extension from the N terminus of pro-σ^K (61).

The discovery of rice etioplast protein 2714.m00157 that potentially regulates transcription by proteolytic activation of transcription factor(s) suggests a novel and unexpected mechanism for dark-specific transcriptional regulation in plastids. Our hypothesis is supported by the recent report of SaSIG3 (*Sinapis alba* L., σ-like factor 3), which has an N-terminal extension reminiscent of the N-terminal domain of pro-σ^K from *B. subtilis* (64). The N-terminal domain of SaSIG3 was inhibitory for productive transcription initiation but not for promoter binding (64, 65). SaSIG3 accumulates to higher levels in dark-grown compared with light-grown mustard seedlings and was found in two forms, the full-length and a truncated protein (64). Together our proteome analysis supports a model that an etioplast-specific membrane-embedded S2P-like metalloprotease regulates gene expression by proteolytic activation of σ-like factor-dependent transcription in plastids.

Conclusions—We have developed a new protocol for the purification of rice etioplasts and demonstrated the purity of the isolated plastids by comparative semiquantitative protein profiling including different cell organelles from the same plant material. Information about the relative abundance of proteins throughout the fractionation procedure allowed us to clearly distinguish etioplast-localized proteins from potential contaminants. This extensive analysis resulted in a high confidence list of etioplast proteins and sets a new quality standard for the proteome analysis of subcellular compartments. The high confidence in the etioplast localization of the iden-

tified proteins allowed us to report new protein functions in the plastid that were unknown until now. A comparison of the etioplast proteome with proteomes of other plastid types revealed etioplast-specific metabolic functions and potential novel mechanisms in the regulation of plastid gene expression. Together our analysis significantly extends the high quality list of new plastid proteins that will facilitate further analysis of plastid biogenesis and function as well as transit peptide and plastid proteome evolution.

REFERENCES

1. Martin, W., and Herrmann, R. G. (1998) Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118,** 9–17
2. Vothknecht, U. C., and Westhoff, P. (2001) Biogenesis and origin of thylakoid membranes. *Biochim. Biophys. Acta* **1541,** 91–101
3. Neuhaus, H. E., and Emes, M. J. (2000) Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51,** 111–140
4. Gray, J. C., Sullivan, J. A., Wang, J. H., Jerome, C. A., and MacLean, D. (2003) Coordination of plastid and nuclear gene expression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358,** 135–145
5. Jarvis, P. (2003) Intracellular signalling: the language of the chloroplast. *Curr. Biol.* **13,** R314–R316
6. Peltier, J. B., Friso, G., Kalume, D. E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K. J. (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* **12,** 319–341
7. Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002) Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11487–11492
8. Peltier, J. B., Emanuelsson, O., Kalume, D. E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D. A., Soderberg, L., Roepstorff, P., von Heijne, G., and van Wijk, K. J. (2002) Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell* **14,** 211–236
9. Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schroder, W. P., and Kieselbach, T. (2002) Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. *J. Biol. Chem.* **277,** 8354–8365
10. Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., and Rolland, N. (2003) Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. *Mol. Cell. Proteomics* **2,** 325–345
11. Froehlich, J. E., Wilkerson, C. G., Ray, W. K., McAndrew, R. S., Osteryoung, K. W., Gage, D. A., and Phinney, B. S. (2003) Proteomic study of the Arabidopsis thaliana chloroplastic envelope membrane utilizing alternatives to traditional two-dimensional electrophoresis. *J. Proteome Res.* **2,** 413–425
12. Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., and Baginsky, S. (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14,** 354–362
13. Lonosky, P. M., Zhang, X., Honavar, V. G., Dobbs, D. L., Fu, A., and Rodermel, S. R. (2004) A proteomic analysis of maize chloroplast biogenesis. *Plant Physiol.* **134,** 560–574
14. Peltier, J. B., Ytterberg, A. J., Sun, Q., and van Wijk, K. J. (2004) New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast, and versatile fractionation strategy. *J. Biol. Chem.* **279,** 49367–49383
15. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858
16. Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A. (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell. Proteomics* **3,** 531–533
17. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
18. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) The sequence of the human genome. *Science* **291,** 1304–1351
19. Karplus, K., Sjolander, K., Barrett, C., Cline, M., Haussler, D., Hughey, R., Holm, L., and Sander, C. (1997) Predicting protein structure using hidden Markov models. *Proteins* Suppl. 1, 134–139
20. Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4,** 117
21. Dunkley, T. P., Watson, R., Griffin, J. L., Dupree, P., and Lilley, K. S. (2004) Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* **3,** 1128–1134
22. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300,** 1005–1016
23. Wendel, J. F., Goodman, M. M., Stuber, C. W., and Beckett, J. B. (1988) New isozyme systems for maize (Zea mays L.): aconitate hydratase, adenylate kinase, NADH dehydrogenase, and shikimate dehydrogenase. *Biochem. Genet.* **26,** 421–445
24. Hodges, M. (2002) Enzyme redundancy and the importance of 2-oxoglutarate in plant ammonium assimilation. *J. Exp. Bot.* **53,** 905–916
25. Baginsky, S., Siddique, A., and Gruissem, W. (2004) Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids. *J. Proteome Res.* **3,** 1128–1137
26. Schleiff, E., and Klosgen, R. B. (2001) Without a little help from 'my' friends: direct insertion of proteins into chloroplast membranes? *Biochim. Biophys. Acta* **1541,** 22–33
27. Miras, S., Salvi, D., Ferro, M., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002) Non-canonical transit peptide for import into the chloroplast. *J. Biol. Chem.* **277,** 47770–47778
28. Silva-Filho, M. C. (2003) One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.* **6,** 589–595
29. Kiessling, J., Martin, A., Gremillon, L., Rensing, S. A., Nick, P., Sarnighausen, E., Decker, E. L., and Reski, R. (2004) Dual targeting of plastid division protein FtsZ to chloroplasts and the cytoplasm. *EMBO Rep.* **5,** 889–894
30. Wall, M. K., Mitchenall, L. A., and Maxwell, A. (2004) Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7821–7826

31. Peeters, N., and Small, I. (2001) Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* **1541,** 54–63

32. Lahiri, S. D., and Allison, L. A. (2000) Complementary expression of two plastid-localized σ-like factors in maize. *Plant Physiol.* **123,** 883–894

33. Flugge, U. I. (1999) Phosphate translocators in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 27–45

34. Neuhaus, H. E., Batz, O., Thom, E., and Scheibe, R. (1993) Purification of highly intact plastids from various heterotrophic plant tissues: analysis of enzymic equipment and precursor dependency for starch biosynthesis. *Biochem. J.* **296,** 395–401

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

36. Schurmann, P. (2003) Redox signaling in the chloroplast: the ferredoxin/thioredoxin system. *Antioxid. Redox Signal.* **5,** 69–78

37. Sakai, A., Kawano, S., and Kuroiwa, T. (1992) Conversion of proplastids to amyloplasts in tobacco cultured cells is accompanied by changes in the transcriptional activities of plastid genes. *Plant Physiol.* **100,** 1062–1066

38. Quick, W. P., Scheibe, R., and Neuhaus, H. E. (1995) Induction of hexose-phosphate translocator activity in spinach chloroplasts. *Plant Physiol.* **109,** 113–121

39. Knight, J. S., and Gray, J. C. (1995) The N-terminal hydrophobic region of the mature phosphate translocator is sufficient for targeting to the chloroplast inner envelope membrane. *Plant Cell* **7,** 1421–1432

40. Niittyla, T., Messerli, G., Trevisan, M., Chen, J., Smith, A. M., and Zeeman, S. C. (2004) A previously unknown maltose transporter essential for starch degradation in leaves. *Science* **303,** 87–89

41. Corbin, R. W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C. E., Jr., Root, K., McAuliffe, J., Jordan, M. I., Kustu, S., Soupene, E., and Hunt, D. F. (2003) Toward a protein profile of Escherichia coli: comparison to its transcription profile. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9232–9237

42. Apel, K., Santel, H. J., Redlinger, T. E., and Falk, H. (1980) The protochlorophyllide holochrome of barley (Hordeum vulgare L.). Isolation and characterization of the NADPH:protochlorophyllide oxidoreductase. *Eur. J. Biochem.* **111,** 251–258

43. Oliver, R. P., and Griffiths, W. T. (1981) Covalent labelling of the NADPH: protochlorophyllide oxidoreductase from etioplast membranes with [^{3}H]N-phenylmaleimide. *Biochem. J.* **195,** 93–101

44. Johanningmeier, U., and Howell, S. H. (1984) Regulation of light-harvesting chlorophyll-binding protein mRNA accumulation in *Chlamydomonas reinhardi*. Possible involvement of chlorophyll synthesis precursors. *J. Biol. Chem.* **259,** 13541–13549

45. Papenbrock, J., and Grimm, B. (2001) Regulatory network of tetrapyrrole biosynthesis—studies of intracellular signalling involved in metabolic and developmental control of plastids. *Planta* **213,** 667–681

46. Rodermel, S. (2001) Pathways of plastid-to-nucleus signaling. *Trends Plant Sci.* **6,** 471–478

47. Strand, A., Asami, T., Alonso, J., Ecker, J. R., and Chory, J. (2003) Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature* **421,** 79–83

48. Kessler, F., and Schnell, D. J. (2004) Chloroplast protein import: solve the GTPase riddle for entry. *Trends Cell Biol.* **14,** 334–338

49. Soll, J., and Schleiff, E. (2004) Protein import into chloroplasts. *Nat. Rev. Mol. Cell. Biol.* **5,** 198–208

50. Ko, K., Taylor, D., Argenton, P., Innes, J., Pedram, B., Seibert, F., Granell, A., and Ko, Z. (2005) Evidence that the plastid translocon Tic40 components possess modulating capabilities. *J. Biol. Chem.* **280,** 215–224

51. Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., and van 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

52. Baena-Gonzalez, E., and Aro, E. M. (2002) Biogenesis, assembly and turnover of photosystem II units. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357,** 1451–1460

53. Baker, M. E., Grundy, W. N., and Elkan, C. P. (1998) Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. *Biochem. Biophys. Res. Commun.* **248,** 250–254

54. Bollenbach, T. J., Schuster, G., and Stern, D. B. (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. *Prog. Nucleic Acids Res. Mol. Biol.* **78,** 305–337

55. Bollenbach, T. J., and Stern, D. B. (2003) Divalent metal-dependent catalysis and cleavage specificity of CSP41, a chloroplast endoribonuclease belonging to the short chain dehydrogenase/reductase superfamily. *Nucleic Acids Res.* **31,** 4317–4325

56. Yang, J., and Stern, D. B. (1997) The spinach chloroplast endoribonuclease CSP41 cleaves the 3′-untranslated region of petD mRNA primarily within its terminal stem-loop structure. *J. Biol. Chem.* **272,** 12874–12880

57. Bollenbach, T. J., Tatman, D. A., and Stern, D. B. (2003) CSP41a, a multifunctional RNA-binding protein, initiates mRNA turnover in tobacco chloroplasts. *Plant J.* **36,** 842–852

58. Hayes, R., Kudla, J., Schuster, G., Gabay, L., Maliga, P., and Gruissem, W. (1996) Chloroplast mRNA 3′-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. *EMBO J.* **15,** 1132–1141

59. Baginsky, S., Shteiman-Kotler, A., Liveanu, V., Yehudai-Resheff, S., Bellaoui, M., Settlage, R. E., Shabanowitz, J., Hunt, D. F., Schuster, G., and Gruissem, W. (2001) Chloroplast PNPase exists as a homo-multimer enzyme complex that is distinct from the Escherichia coli degradosome. *RNA* **7,** 1464–1475

60. Stern, D. B., and Gruissem, W. (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* **51,** 1145–1157

61. Rudner, D. Z., Fawcett, P., and Losick, R. (1999) A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14765–14770

62. Lewis, A. P., and Thomas, P. J. (1999) A novel clan of zinc metallopeptidases with possible intramembrane cleavage properties. *Protein Sci.* **8,** 439–442

63. Zelenski, N. G., Rawson, R. B., Brown, M. S., and Goldstein, J. L. (1999) Membrane topology of S2P, a protein required for intramembranous cleavage of sterol regulatory element-binding proteins. *J. Biol. Chem.* **274,** 21973–21980

64. Homann, A., and Link, G. (2003) DNA-binding and transcription characteristics of three cloned sigma factors from mustard (Sinapis alba L.) suggest overlapping and distinct roles in plastid gene expression. *Eur. J. Biochem.* **270,** 1288–1300

65. Hakimi, M. A., Privat, I., Valay, J. G., and Lerbs-Mache, S. (2000) Evolutionary conservation of C-terminal domains of primary sigma^{70}-type transcription factors between plants and bacteria. *J. Biol. Chem.* **275,** 9215–9221

66. Deleted in proof