Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters

Myriam Ferro*, Daniel Salvi†, Hélène Rivière-Rolland‡§, Thierry Vermat§, Daphné Seigneurin-Berny†, Didier Grunwald¶, Jérôme Garin*, Jacques Joyard†, and Norbert Rolland†‖

*Laboratoire de Chimie des Protéines, Equipe de Recherche Méthodologique 0201, Institut National de la Santé et de la Recherche Médicale/Commissariat à l'Energie Atomique, †Laboratoire de Physiologie Cellulaire Végétale, Unité Mixte de Recherche 5019, Centre National de la Recherche Scientifique/Commissariat à l'Energie Atomique/Université Joseph Fourier, and ¶Laboratoire Canaux Ioniques et Signalisation, Institut National de la Santé et de la Recherche Médicale E9931, Département Réponse et Dynamique Cellulaires, Commissariat à l'Energie Atomique-Grenoble, F-38054 Grenoble Cedex 9, France; ‡Institut National de Recherche en Informatique et en Automatique Rhône-Alpes, 655 Avenue de l'Europe, Montbonnot, F-38334 Saint Ismier Cedex, France; and §GENOME Express, 11 Chemin des Prés, F-38944 Meylan Cedex, France

Communicated by Roland Douce, Université de Grenoble, Grenoble, France, July 2, 2002 (received for review May 10, 2002)

A two-membrane system, or envelope, surrounds plastids. Because of the integration of chloroplast metabolism within the plant cell, the envelope is the site of many specific transport activities. However, only a few proteins involved in the processes of transport across the chloroplast envelope have been identified already at the molecular level. To discover new envelope transporters, we developed a subcellular proteomic approach, which is aimed to identify the most hydrophobic envelope proteins. This strategy combined the use of highly purified and characterized membrane fractions, extraction of the hydrophobic proteins with organic solvents, SDS/PAGE separation, and tandem mass spectrometry analysis. To process the large amount of MS/MS data, a BLAST-based program was developed for searching in protein, expressed sequence tag, and genomic plant databases. Among the 54 identified proteins, 27 were new envelope proteins, with most of them bearing multiple α-helical transmembrane regions and being very likely envelope transporters. The present proteomic study also allowed us to identify common features among the known and newly identified putative envelope inner membrane transporters. These features were used to mine the complete *Arabidopsis* genome and allowed us to establish a virtual plastid envelope integral protein database. Altogether, both proteomic and *in silico* approaches identified more than 50 candidates for the as yet previously uncharacterized plastid envelope transporters. The predictable function of some of these proteins opens up areas of investigation that may lead to a better understanding of the chloroplast metabolism. The present subcellular proteomic approach is amenable to the analysis of the hydrophobic core of other intracellular membrane systems.

Plastids, and especially chloroplasts, conduct vital biosynthetic functions, and many reactions are located exclusively within these unique organelles. A two-membrane system, the envelope, surrounds all plastid types and separates the plastid stroma from the cytosol. As a consequence, the envelope is involved in the controlled exchange of a variety of ions and metabolites between these two subcellular compartments (1).

Chloroplasts import cytoplasmically synthesized precursor proteins from the cytosol. Translocation of precursor proteins across the envelope is achieved by the joint action of Toc and Tic translocons located at the outer and inner envelope membranes, respectively, of the chloroplast envelope (2, 3). Chloroplasts also take up intermediates of various metabolic pathways such as dicarboxylic acids, acetate, and phosphoenolpyruvate. Chloroplasts also have been demonstrated to import inorganic ions like K^+, Na^+, Mg^{2+}, Ca^{2+}, Cl^-, NO_2^{2-}, SO_4^{2-}, PO_4^{2-}, Fe^{2+} (4, 5). As the sole site of biosynthesis of most amino acids (with the exception of sulfur-containing amino acids; refs. 6 and 7), chloroplasts must export these compounds for protein synthesis in the cytosolic and mitochondrial compartments. Finally, because of metabolism compartmentation, several other organic or inorganic compounds (other metals, fatty acids, cofactors, vitamins, etc.) are suspected to cross envelope membranes through as-yet-uncharacterized mechanisms.

Since the identification and cloning of the triose phosphate-3-phosphoglycerate-phosphate translocator (8), few proteins involved in the processes of ion and metabolite transport across the chloroplast envelope have been identified at the molecular level. To date, the sequence of substrate-specific outer membrane solute channels, a family of inner membrane transport proteins catalyzing the movement of phosphorylated intermediates and a new type of ATP/ADP transporter are already available (4).

Previously, it was a significant accomplishment simply to clone an ion transport gene. With the recent completion of the *Arabidopsis* genome-sequencing project (9), hundreds of known and putative transporters have been identified in the *Arabidopsis* genome sequence (see the AMPL database, available at www.cbs.umn.edu/arabidopsis; ref. 10). However, the subcellular localization and *in planta* function of most of these proteins remain to be determined. During the present work, a targeted hydrophobic and subcellular specific proteomic approach was developed to identify components of the hydrophobic core of the chloroplast envelope. We demonstrate that the development of such studies may help to reveal the hydrophobic protein composition of a particular membrane system in a given tissue, thus enhancing our understanding of the plant transport systems.

Methods

Purification of Chloroplast Envelope Membranes. Crude chloroplasts were obtained from spinach (*Spinacia oleracea* L.) leaves and purified by isopycnic centrifugation using Percoll gradients (11). Purified, intact chloroplasts were lysed in hypotonic medium, and envelope membranes were purified from the lysate by centrifugation on sucrose gradients (11). Envelope subfractions enriched in outer and inner membranes were obtained, as already described (12). Chloroplast envelope membrane preparations have been characterized extensively and were demonstrated to be devoid of contamination by extra-plastidial membranes and by thylakoids (11).

Chloroform/Methanol (C/M) Extractions. Chloroplast envelope hydrophobic proteins were extracted from envelope preparations by using a C/M mixture (13, 14). Envelope membranes (0.5 mg

Abbreviations: C/M, chloroform/methanol; GFP, green fluorescent protein; TP, transit peptide; TM, transmembrane; Res, number of amino acid residues.

Data deposition: The sequence reported in this paper has been deposited in the EMBL database (accession no. AJ491150).

‖To whom reprint requests should be addressed. E-mail: nrolland@cea.fr.

PLANT BIOLOGY

of proteins in 0.1 ml of storage buffer) were slowly diluted in 0.9 ml of cold C/M (2:1, vol/vol) solution. The resulting mixture was stored for 15 min on ice before centrifugation (4°C, 20 min, 12,000 × *g*). Proteins insoluble in the organic phase were recovered as a white pellet, and proteins present in the organic phase were analyzed further. Protein contents of membrane fractions were estimated by using the Bio-Rad protein assay reagent (15).

SDS/PAGE and Western Blot Analyses. Proteins present in the C/M extracts were precipitated with acetone, resuspended in 50 μl of SDS/PAGE buffer, and finally loaded on 12% acrylamide gels for SDS/PAGE analyses (16). Rabbit polyclonal antibodies were produced, as previously described (13), against four synthetic peptides, corresponding to residues 532–546 and 483–494 of IEP60 and 221–231 and 462–471 of HP45. Immune sera were purified as described (17). The envelope proteins were detected with the purified antibodies diluted 1:1,000 by using alkaline phosphatase staining. Preimmune sera gave no signal.

Mass Spectrometry and Protein Identification. After separation by SDS/PAGE, discrete bands were excised from the Coomassie blue-stained gel. The in-gel digestion was carried out as described (14). Gel pieces then were extracted with 5% (vol/vol) formic acid solution and acetonitrile. The extracted peptides were desalted by using C18 Zip Tips (Millipore). Elution of the peptides was performed with 5–10 μl of a 50:50:0.1 (vol/vol) acetonitrile/H_2O/formic acid solution. The peptide solution was introduced into a glass capillary (Protana, Odense, Denmark) for nanoelectrospray ionization. Tandem mass spectrometry experiments were carried out on a Q-TOF hybrid mass spectrometer (Micromass). Interpretation of MS/MS spectra was achieved manually and with the help of the PEPSEQ program (MassLynx software, Micromass, Manchester, U.K.). MS/MS sequence information were used for database searching by using a home–made BLAST-based program. The so-called BLASTCOMP program (see Fig. 4, which is published as supporting information on the PNAS web site, www.pnas.org) allows a BLAST search for each amino acid sequence and allows one to cluster amino acid sequences with a common BLAST hit. BLASTP and TBLASTN were used with specific matrices to mine plant protein and genomic databases, respectively.

Prediction Methods. Predictions for chloroplast localization and membrane-spanning regions were achieved by using the software programs CHLOROP (18) and HMMTOP (19), respectively.

Transient Expression of Protein Fusions in *Arabidopsis*. The green fluorescent protein (GFP) reporter plasmid 35Ω-sGFP(S65T) and the plasmid 35Ω-TP-sGFP(S65T) containing the transit peptide (TP) sequence from RBCs fused to GFP were described (20). Construction of the plasmid containing the precursor of the IEP60 spinach protein fused to GFP (35Ω-*SoPht2;1*-sGFP(S65T)) was performed as follows. The complete coding region of IEP60 was PCR-amplified by using two flanking primers, *Xho*I-N-ter TCACTCGAGATGACTTCCTCTTGCCTCTTATC and *Xho*I-C-ter TCACTCGAGTAATACGTAGGATAAACCCTTGG. This fragment was inserted into the *Sal*I-digested plasmid 35Ω-sGFP(S65T). Correct orientation and sequence of the inserted fragments were controlled. Plasmids (5 μg; Qiagen Plasmid MidiKit) were introduced to *Arabidopsis* leaves (3–4 weeks old) by using a pneumatic particle gun (Bio-Rad PDS-1000/He, helium pressure of 1,550 psi, 1,350 psi Rupture Disks, 10-cm target distance, 1-μm gold Microcarriers). After bombardment, leaves were incubated on MS plates for 18–36 h in dim light. Fluorescence microscopy was performed with a confocal laser-scanning microscope (TCS-SP2, Leica, Deerfield, IL; ref. 21).

Results

Isolation of Highly Hydrophobic Chloroplast Envelope Proteins. In total, 54 proteins were identified from 306 nonredundant peptide sequences (Table 1). All known proteins were chloroplastic proteins, thus confirming that no cross contamination from extra-plastidial membranes occurred. No contamination from thylakoids was found. According to their location, four categories of proteins were found: (*i*) inner envelope membrane proteins, (*ii*) outer envelope membrane proteins, (*iii*) peripheral and stroma proteins, and (*iv*) proteins with unknown subcellular localization. Peripheral and stroma proteins are obviously soluble contaminants, because these contaminants could not be visualized on the stained SDS/PAGE gel (e.g., Rbcl; ref. 14). On the other hand, highly hydrophobic proteins were found to be highly enriched in the same SDS/PAGE gel (14). Because of this enrichment and the mass spectrometry sensitivity, it was possible to identify very low abundant, highly hydrophobic proteins. Indeed, we could identify P60, a highly hydrophobic protein [13 predicted transmembrane (TM) domains] representing less than 3% of the C/M soluble envelope proteins (1:100,000th of total cellular proteins).

Most (80%) of the identified proteins contain at least 1 predicted TM domain (see Fig. 5, which is published as supporting information on the PNAS web site). The proteins that did not contain any predicted TM domain could be classified in three categories: (*i*) genuine integral membrane proteins that contain amphipatic β-strands spanning the membrane (e.g., OEP21; ref. 22); (*ii*) peripheral proteins and stroma soluble contaminants; and (*iii*) genuine envelope proteins that are extracted in C/M solutions because of strong interactions with lipids (Monogalactosyldiacylglycerol synthase; ref. 23). Twenty-one proteins were found to have at least four predicted α-helical TM domains. Because no contamination from other membrane systems has been detected, these proteins are part of the highly hydrophobic core of the chloroplast envelope.

Identification of New Putative Transport Systems. Except for β-strands containing proteins such as porins, most proteins involved in transport functions bear at least four α-helical TM domains (see AMPL database). Of the 21 identified proteins that exhibit at least 4 predicted TM domains, only 6 proteins have a fully characterized function, and 4 are known to be chloroplast envelope transporters (Table 1). Among the 17 remaining proteins, 10 have a close or remote similarity with characterized transport systems, 1 shares a low similarity with a transferase, and 6 are not related to any known protein. Therefore, regarding the high purity of the chloroplast envelope fraction, the present subproteomics approach allowed us to identify several new chloroplast envelope gene products that are very likely transporters.

Validation of the Subcellular Location of Newly Identified Highly Hydrophobic Proteins. Although the present subproteomics strategy was highly specific to the subcellular location, the location of new proteins remained to be validated. As controls, two of the most minor highly hydrophobic proteins, P60 and P45, were considered for immunolocalization experiments. As shown in Fig. 1*A*, both proteins were immuno-detected in the envelope fraction. No reaction was detected in the chloroplast extract, stroma, or thylakoid subfractions. Further experiments demonstrated that P60 and P45 are, in fact, associated with the inner membrane of the chloroplast envelope (Fig. 1*B*) and, thus, were called IEP60 and IEP45.

The complete sequence of the spinach IEP60 cDNA was obtained by using rapid amplification of cDNA ends (RACE)/PCR experiments. If one excludes the 95 *N*-ter amino acids, the spinach IEP60 amino acid sequence is almost identical (82% identity, 88% similarity) to the *Arabidopsis Pht2;1* H^+/Pi trans-

Table 1. Features of the identified chloroplast envelope proteins

Da	Acc nb	Species	Protein names and AtDB annotations	cTP	TM	R/TM	pI	Loc	AGI Acc nb
60,288	Q41364	So	IEP45= 2-oxoglutarate/malate translocator	+ (70)	14	41	9.7	IM	At5g12860
57,616	Q9LLE2	So	IEP62 (putative sugar transporter)	+ (85)	12	46	9.3	IM	At5g16150
44,234	P11869	So	IEP30= Triose-P/Pi translocator	+ (80)	8	50	9.7	IM	At5g46110
45,469	Q9MSB5	Mc	IEP33 = Pi/PEP translocator	+ (89)	6	69	10.0	IM	At5g33320
37,997	Q9LLY4	Bn	LPAAT = lysophosphatidic acid acyl transferase	+ (87)	4	86	9.7	IM #	At4g30580
51,307	P48629	So	FD6C= omega-6 fatty acid desaturase	+ (64)	4	112	9.4	IM #	At4g30950
38,976	P23525	So	IEP37= SAM-dependent methyl transferase	+ (62)	1	344	9.5	IM	At3g63410
57,511	Q9SM44	So	MGDS =monogalactosyldiacylglycerol synthase	+ (64)	0	-	9.2	IM	At4g31780
54,017	Q9FVQ4	At	HP45 unknown function	+ (13)	12	43	9.6	IM *	At1g32080
60,198	cloning*	So	IEP60 = phosphate transporter	-	13	44	9.1	IM *	At3g26570
22,911	O82251	At	IEP16 (low homology TIC20)	+ (49)	4	52	10.3	IM *	At2g47840
25,234	cloning*	So	IEP18 unknown function	+ (58)	4	59	9.8	IM *	At5g62720
43,088	AX046732	Gm	HP43 (low homology UBIA prenyltransferase)	+ (87)	9	43	9.9	IM ?	At3g11950
59,829	Q9FIF2	At	HP59 (D-xylose-H^+ symporter - like protein)	+ (31)	12	46	9.0	IM ?	At5g59250
59,991	Q9FMF7	At	HPSOT (2-oxoglut/malate translocator homologue)	+ (68)	12	47	9.3	IM ?	At5g64290
36,066	O81017	At	HP36 (putative Na^+/taurocholate transporter)	-	7	48	8.9	IM ?	At2g26900
34,874	Q9SYM0	At	HP34 unknown function	+ (65)	6	55	9.8	IM ?	At1g78620
62,937	P92935	At	HPTLC (similar to adenine nucleotide translocase)	+ (66)	10	57	10.0	IM ?	At1g15500
26,992	Q9SJH9	At	HP25 (similar to 22 kDa perox membrane protein)	+ (46)	4	61	9.7	IM ?	At2g42770
27,986	Q9SD32	At	HP28 unknown function	+ (35)	4	62	10.2	IM ?	At3g51140
29,585	Q9M277	At	HP29b unknown function	+ (50)	4	68	9.3	IM ?	At3g61870
41,323	CAC42908	At	HP40 unknown function	+ (61)	5	77	7.0	IM ?	At5g12470
28,911	Q9FNA1	At	HP29c unknown function	+ (60)	3	87	9.0	IM ?	At5g13720
35,663	Q9SVB2	At	HP35 (putative mitochondrial carrier-like protein)	+ (23)	3	110	9.6	IM ?	At4g39460
17,829	Q9C7S3	At	HP17 unknown function	+ (59)	1	168	8.6	IM ?	At1g42960
41,817	O65023	At	HP42 (put perox Ca^{2+}-dependent solute carrier)	+ (26)	2	190	9.8	IM ?	At3g51870
52,800	Q9SHK7	At	HP52 (aldo/keto reductase-like protein)	+ (31)	2	238	9.4	IM ?	At1g06690
38,841	Q9LGI8	Os	HP32 unknown function	+ (68)	1	294	9.1	IM ?	At4g23430
41,630	Q9LTR2	At	HP41b unknown function	+ (45)	1	381	8.9	IM ?	At3g20320
44,552	O80565	At	HP44 unknown function	+ (73)	1	391	9.1	IM ?	At2g43950
30,105	Q9SZB2	At	HP27 (TIC22-like protein)	+ (59)	0	-	9.3	IM ?	At4g33350
6,479	P19407	So	OEP6 unknown function	-	1	62	6.1	OM	At3g63160
16,294	S38569	So	OMP24 unknown function	-	1	148	4.8	OM	At3g52230
34, 723	Q9GDD2	At	TOC34 (protein import apparatus component)	-	1	313	9.4	OM	At5g05000
43,359	P21218	At	POR = protochlorophyllide reductase (A or B)	+ (43)	1	401	9.2	OM+TB	At4g27440
89,189	Q9STE8	At	OEP75 = chloroplast import-associated channel	+ (79)	1	818	8.9	OM	At3g46740
20,444	Q9SM57	Ps	OEP21 = ATP-regulated solute channel	-	0	-	9.5	OM	At1g76405
27,982	Q9SCK3	At	HP30 (low homology TIM17/TIM22)	-	4	65	9.5	E ?	At3g49560
21,666	Q9SZ09	At	HP20 (low homology TIM17/TIM22)	-	3	69	7.6	E ?	At4g26670
49,848	AAK83606	At	HP50 unknown function	-	1	451	8.8	E ?	At2g44640
49,190	O80503	At	HP49 unknown function	+ (99)	0	-	9.0	E ?	At2g44640
32,356	Q9S834	At	NCLPP1 (put. ATP-dependent CLP protease)	+ (62)	2	149	8.4	S / EB	At1g02560
31,208	Q9XJ36	At	NCLPP2 (put. ATP-dependent CLP protease)	+ (54)	1	279	9.2	S / EB	At1g12410
31,572	Q42564	At	AP = Ascorbate peroxidase	-	1	287	6.5	S / TB	At4g35000
80,281	O20250	So	TKTC= Transketolase	+ (65)	2	370	6.2	S / TB	At3g60750
41,189	P00455	So	FENR = Ferredoxin-$NADP^+$ reductase	+ (54)	0	-	8.7	S / TB	At5g66190
42,468	P16096	So	ALFC = Fructose-bisphosphate aldolase	+ (45)	1	394	6.9	S	At4g38970
26,562	Q9M5A8	Le	CH1C (chloroplast Cpn21 chaperonin)	+ (50)	0	-	6.9	S ?	At5g20720
34,569	P16016	So	CAHC= Carbonic anhydrase	+ (64)	0	-	6.6	S	At3g01500
52,740	P00875	So	RBL = Rubisco large subunit	-	0	-	6.1	S	Plastid gene
45,573	P29409	So	PGKH = Phosphoglycerate kinase	+ (27)	0	-	5.8	S	At3g12780
51,485	P10871	So	RCA = Rubisco activase	+ (57)	0	-	6.3	S	At2g39730
36,226	P19866	So	G3PA = Glyceraldehyde 3-P dehydrogenase A	-	0	-	6.7	S	At1g12900

Da, molecular weight (protein precursor); Acc nb, accession number in SwissProt, TrEMBL, or NCBI; AGI acc nb, AGI accession numbers; At, *Arabidopsis thaliana*; So, *Spinacia oleracea*; Mc, *Mesenbryanthenum crystallinum*; Gm, Glycine max; Le, *Lycopersicon esculentum*; Os, *Orysa sativa*; Bn, *Brassica napus*; cTp, predicted plastid transit peptide and maturation site (18); TM, predicted TM domains (19); R, number of residues of protein precursor; Loc, subplastidial location; E, envelope; IM, inner membrane; OM, outer membrane; S, stroma; EB or TB, referenced interaction with envelope or thylakoid membranes, # based on localization of enzyme activity (LPAAT, ref. 12; FD6C, ref. 21).
*Localization or cloning during this work.

porter (ref. 24, and see Figs. 6 and 7, which are published as supporting information on the PNAS web site). Therefore, it can be assumed that these proteins are orthologous. Because the *Arabidopsis Pht2;1* Pi transporter was previously suggested to be localized in the plasma membrane and involved in the uptake and intercellular movement of Pi in shoot organs (24, 25), the question of whether IEP60 is exclusively localized within the chloroplast envelope was addressed. Therefore, transient expression of spinach IEP60 fused to the GFP was performed in *Arabidopsis* leaves. Two control experiments were included: transient expression (*i*) of GFP alone, which was targeted to the cytosol, and (*ii*) of GFP fused to the transit sequence of the RuBisCO small subunit (TP-GFP), which was targeted to the chloroplasts of transformed cells (Fig. 1*C*). Fluorescence of

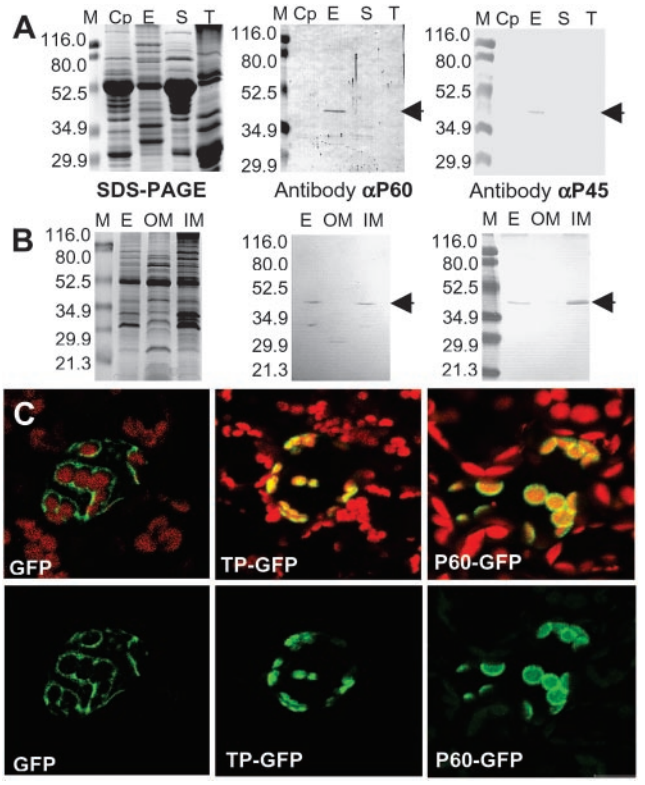

Fig. 1. Subcellular localization of the IEP60 (*Pht2;1* phosphate transporter) and HP45 proteins in (*A*) the chloroplast envelope and (*B*) the inner membrane of the chloroplast envelope. M, markers; Cp, crude chloroplast extract; S, stroma; T, thylakoid; E, envelope; IM, inner membrane; OM, outer membrane (20 μg of proteins per lane). SDS/PAGE and Western blots performed with the antibodies raised against the IEP60 and HP45 synthetic peptides. (*C*) Plastid targeting of IEP60 in *Arabidopsis* leaves. Images correspond to the superimposition of 8–10 optical sections, each 1 μm thick. GFP, 35Ω-sGFP(S65T) plasmid. TP-GFP, 35Ω-TP-sGFP(S65T) plasmid containing the TP from RBCs fused to GFP. P60-GFP, 35Ω-*SoPht2;1*-sGFP(S65T) plasmid containing the precursor of the spinach IEP60 protein fused to GFP.

IEP60::GFP fusions seems to be more concentrated to the periphery of the chloroplasts. No fluorescence was detected at the periphery of the transformed cells (Fig. 1*C*), thus suggesting that the proposed localization of the spinach P60 in the plasma membrane has to be excluded. The spinach IEP60 H^+/Pi transporter is, therefore, exclusively located in the inner membrane of the chloroplast envelope.

Chloroplast Envelope Transport Systems Share Common Features. In chloroplasts, active transport functions are located in the inner envelope membrane. Consequently, a protein located in the inner membrane has a greater probability to be a transporter if it is highly hydrophobic. Indeed, the few known transporters of the chloroplast envelope (IEP30, IEP45, IEP33, IEP62, and IEP60) are known to be located in the inner membrane and contain at least four TM α-helices. Their strong hydrophobicity is also revealed by a low number of amino acid residues (Res)/TM value, below 100. Interestingly, these known transporters and other proteins of the inner envelope membrane that were identified during the course of this work exhibit a very high pI (Table 1). Other identified proteins with putative transport function were shown to meet these criteria. To determine whether any correlation exists between the location of the proteins and their physico-chemical properties, the pI and Res/TM values were calculated for each protein listed in Table 1. As shown in Fig. 2, a strong correlation was found between the

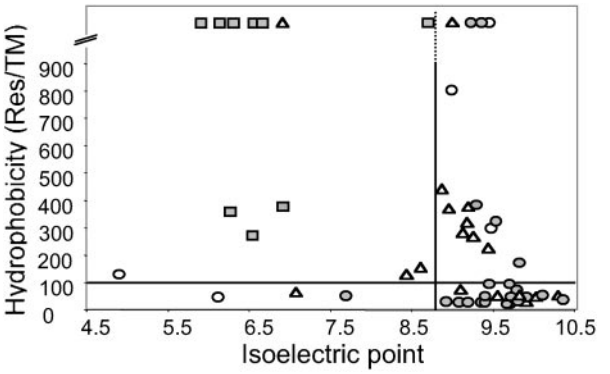

Fig. 2. Identification of common features in the proteins identified by means of the proteomic approach. Position of the spots was determined according to hydrophobicity (Res/TM ratio), and calculated pI of all proteins was identified through the proteomic approach (Table 1). Inner envelope proteins (gray circles), outer envelope proteins (white circles), stroma proteins (squares), and proteins of unknown subplastidial localization (triangles).

location and combined values of pI and Res/TM, especially for the inner membrane proteins. Indeed, only proteins from the inner membrane were found to have both a Res/TM < 100 and pI > 8.8. Knowing that (*i*) active transport functions are generally located in the inner membrane, (*ii*) transport systems are highly hydrophobic proteins (TM ≥ 4), and (*iii*) the present proteomic approach showed that chloroplast envelope proteins having Res/TM < 100 and pI > 8.8 are very likely to be located in the inner membrane, we could assume that chloroplast envelope proteins sharing such properties are very likely to be inner envelope transport systems. Among the proteins identified in the present study and showing these features, eight proteins are homologous to known transporters from various species (IEP62, HP59, HPSOT2, HP36, IEP16, HP25, and HP30), and five have unknown function and location (HP45, HP34, IEP18, HP28, and HP29c). These proteins are, therefore, good candidates for further functional analysis of transport functions in the chloroplast envelope.

Proteomics Feeds Bioinformatics for Database Mining. The list of previously uncharacterized putative transport systems of the chloroplast envelope is certainly not exhaustive. Nevertheless, the present proteomic approach gave us keys to search for additional putative chloroplast envelope transporters in plant databases. As a matter of fact, all chloroplast proteins have a chloroplast TP, except for some outer envelope membrane proteins and for proteins coded by the chloroplast genome. It should be noted that for TP prediction, ChloroP (18) was found to be the only reliable tool for prediction of TP in inner envelope protein sequences (Table 1). Furthermore, the present study showed that envelope transporters have specific pI, TM, and Res/TM range of values. Some other chloroplast proteins that are located in the thylakoid membranes are highly hydrophobic and have a chloroplast TP. However, database observation showed that thylakoid membrane proteins are generally more acidic proteins. Thus, apart from few exceptions, the combination of (*i*) the prediction of a chloroplast TP, (*ii*) Res/TM < 100, (*iii*) TM ≥ 4, and (*iv*) pI > 8.8 is characteristic to chloroplast envelope transporters (see Table 1). As these parameters can be predicted by bioinformatic tools from polypeptide sequences, they were used to mine the AMPL database (Fig. 3).

By using this *in silico* approach, only 136 proteins (of the 25,498 predicted *Arabidopsis* proteins) were pointed out. About 35% of these proteins correspond to proteins belonging to plant transporter families (see Table 2, which is published as supporting information on the PNAS web site) or to proteins homolo-

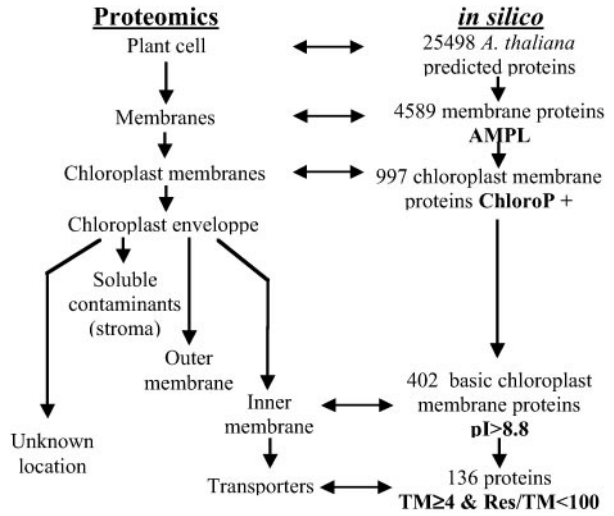

Fig. 3. Combination of proteomics and *in silico* analyses to identify transporters from the inner membrane of the chloroplast envelope. The proteomics approach showed that (*i*) CHLOROP (18) is the most reliable program for the prediction of chloroplast envelope TP prediction, (*ii*) known envelope inner membrane proteins have specifically pI > 8.8, and (*iii*) known envelope inner membrane transporters have TM ≥ 4 and Res/TM < 100. These parameters were used to screen the AMPL database (10).

gous to transport systems present in other species (Pi, PEP, glucose-6-P, ATP/ADP, Na^+-dependent, Na^+/H^+, proteins, C4-dicarboxylates, sugars, vitamins, amino acids, etc.). A few percent of these proteins are proteins involved in lipid or pigment metabolisms (see Table 3, which is published as supporting information on the PNAS web site), providing further evidence for a role of chloroplast envelope membranes in lipid and pigment metabolism (1). The remaining 50% of these proteins are hypothetical proteins whose function could not be predicted on the basis of their primary structure. About 15% of these 136 proteins correspond to proteins identified by means of the proteomic approach; these were essentially proteins involved in validated or putative transport functions (see Tables 2 and 3).

Discussion

The present subcellular proteomic study allowed identification of more than 50 chloroplast proteins (Table 1), and most of them are genuine envelope membrane proteins. However, it should be noted that C/M mixtures extracted more proteins. Indeed, less than 60% of the tryptic peptides could be assigned to proteins leaving 125 orphan peptides that potentially belong to other proteins. On the one hand, identifications were achieved by similarity searching. Thus, for spinach peptides that are not highly similar to *A. thaliana* or other plant sequences, protein identification cannot be successful. On the other hand, by using TBLASTN-based BLASTCOMP, we could point out several genomic regions for which no corresponding expressed sequence tag or predicted protein exists in the *A. thaliana* genome, thus suggesting that some actual proteins are not predicted in the present *A. thaliana* genome annotation (data not shown). Such a proteomic approach provides further evidence that present genome annotations are just predictions that must be confirmed by analyses of the actual gene products (cDNAs or proteins).

Respective Advantages of the Proteomic and *in Silico* Approaches. The present proteomic approach gave us criteria to investigate plant protein databases to identify new chloroplast transport systems *in silico*. According to these criteria, almost 50 proteins (see Table 2) were predicted to be highly basic, highly hydrophobic, located in the chloroplast, and homologous to transport systems, thus being good candidates as transporters of the plastid inner envelope. Bioinformatic tools are the necessary complement of experimental approaches and are of great help in orientating a functional study. Whereas proteomic analyses of chloroplast envelope membranes is likely to identify proteins present in this given membrane system at a precise stage of development, the *in silico* approach also may provide identification of plastid proteins restricted to other plastid types (proplastids, etioplasts, amyloplasts, leucoplasts, and chromoplasts) or to given stages of development, proteins that would, therefore, not be present in chloroplast envelope membranes from mature leaves.

Nevertheless, referring back to our proteomic approach, we suspect that many transporters are missed by the *in silico* study for several reasons. First, most outer envelope transport systems (porins or components of the protein import machinery) do not contain predictable TP sequences or TM α-helices and, therefore, escape to the *in silico* selection (Tables 1 and 2). Second, the CHLOROP program is not 100% reliable. As a matter of fact, *Arabidopsis* IEP30 and spinach IEP60 proteins, for example—which are actually located in the inner membrane of the chloroplast envelope—are not predicted to be chloroplastic (Tables 1 and 2). Third, some chloroplast TPs cannot be detected because of erroneous predictions of protein *N*-ter. For example, although not predicted to be located in the chloroplast by the *in silico* approach, the HP36 protein was identified in the chloroplast envelope during the subcellular proteomic study (Table 1). Further 5′ RACE-PCR experiments demonstrated that the predicted HP36 *N*-ter was, in fact, truncated, and that the corrected protein sequence exhibits a predictable TP (not shown) as the four other members of this protein family (see Table 2). Thus, as demonstrated in the present work, a subcellular proteomic study is mandatory to assign an intracellular localization to proteins lacking classical subcellular targeting signals. It also can detect errors that sometimes occur during genome annotation, such as *in silico* predictions of gene structures.

Nature of the Identified Plastid Envelope Transporters. The combined proteomic and *in silico* approaches allowed us to suggest that several new known or putative transport systems may be associated to the chloroplast envelope. These identified proteins can be classified according to the following groupings: (*i*) proteins of known function already localized in the envelope (e.g., Triose-P/Pi translocator); (*ii*) proteins of known function previously mislocalized (e.g., IEP60 H^+/Pi transporter); (*iii*) expected proteins of predictable function that were not localized (e.g., HPTLC ATP/ADP translocator homologue, SO_4^{2-} or folate transporters); (*iv*) unexpected proteins of predictable function (e.g., IEP60 H^+/Pi or taurocholate transporters); and (*v*) proteins of unpredictable function (HP45, HP34, etc.).

The proteomic approach detected proteins of these five groups. For example, although members of the triose-P/Pi, PEP/Pi, or Glucose-6P/Pi translocators were previously localized in the chloroplast envelope (4), the H^+/Pi transporter was reported to be localized in the plasma membrane (24, 25). Subcellular localization of this ion transporter now raises the question of the *in planta* function of this chloroplast envelope Pi transport system. All previously known envelope Pi translocators catalyze an equimolar exchange of Pi. In creating a net import of Pi in the chloroplast, the H^+/Pi transporter may be essential to maintain the stromal Pi concentration required to initiate the Calvin cycle (26). Finally, the presence of this H^+/Pi transporter raises the question of the origin of the pH gradient across the inner membrane of the chloroplast envelope, because the envelope ATPase (27) still remains to be identified.

Concomitant proteomic identification of two members of the

2-oxoglutarate/malate translocator family (Table 1) suggests that these two proteins may not be differentially expressed spatially or temporally but may differ in substrate specificity: the HPSOT2 protein may catalyze the transport of different dicarboxylic acids (oxaloacetate/malate, malate/glutamate, glutamate/glutamine). Similarly, identification of a close homologue (by HPTLC protein, see Table 1) of the nongreen plastid ATP/ADP translocator identified in proplastids (Table 1) allows one to postulate that expression of this new member of the family may be restricted to chloroplasts.

The presence of sugar transporters like the glucose-6P/Pi translocator (28) and a putative glucose transporter (29, 30) was already demonstrated in chloroplast envelope membranes. Several previously uncharacterized sugar transporters were identified during this study (Tables 1 and 2). Consequently, the glucose transport activity measured on intact chloroplasts—which was solely attributed to the IE62 putative sugar transporter (30)—could result from the concomitant activity of several proteins (at least IEP62 and HP59, Tables 1 and 2).

The identification of several putative Na^+-dependent taurocholate transporters (five members identified) also raises several questions. Because bile acids such as taurocholate do not occur naturally in plants, the nature of the transported compound remains to be determined (31). Such a question was previously addressed because plant ATP-dependent taurocholate transport activity was already detected on vacuolar membranes (32). The identification of these proteins also raises the question of the nature of the Na^+ gradient required to energize this transport.

The identification of several other proteins is consistent with transport activities already associated with the chloroplast envelope. For example, of the twelve potential SO_4^{2-} transporters present in the AMPL database, only one member was identified during this study. As chloroplasts are the sole site of sulfate reduction (33), and because SO_4^{2-} transport across the envelope was demonstrated (34), this protein is a good candidate to catalyze this uptake of SO_4^{2-} into the stroma.

Although many amino acid transporters were identified in plants (35), the nature of the protein that drives the export of these compounds from their unique site of synthesis (the chloroplast) to the cytosol remains to be identified. Identification of members of the amino acid transporter families during this study provides candidates that may catalyze this transport activity.

Because of metabolism compartmentation, several other organic or inorganic compounds are suspected to cross the plastid envelope membranes through as-yet-uncharacterized mechanisms. For example, although the mitochondria were demonstrated to be the sole site of dihydrofolate synthesis in the plant cell, folate-mediated reactions were identified in the cytosol, the mitochondria, and the plastids (36), thus suggesting that folate must be imported in the chloroplast. Identification of a protein highly homologous to animal folate transporters (Table 2) may validate this hypothesis.

All of the functions mentioned above need to be validated by relevant functional studies. The task promises to be difficult, especially for proteins belonging to the same transporter family, as the expression of other gene products with redundant function may lead to uninformative reverse genetic experiments. In addition to putative transporters, both proteomic and *in silico* studies pointed out gene products without any functional homology and which could be correlated potentially to transport activities associated with chloroplast envelope. We believe that the present work opens up new perspectives for a better understanding of plastid metabolism.

We thank Dr. Alain Viari, Institut National de Recherche en Informatique et en Automatique Rhône-Alpes, Montbonnot, France, for original discussions concerning the development of the BLASTCOMP program. We thank Samantha Warren for her critical reading of the manuscript. This work was supported by Génoplante Grant NO19993663 and Centre National de la Recherche Scientifique and Commissariat à l'Energie Atomique research programs.

1. Joyard, J., Teyssier, E., Miege, C, Berny-Seigneurin, D, Maréchal, E., Block, M. A., Dorne, A. J., Rolland, N., Ajlani, G. & Douce, R. (1998) *Plant Physiol.* **118,** 715–723.
2. Jarvis, P. & Soll, J. (2001) *Biochim. Biophys. Acta* **1541,** 64–79.
3. Jackson-Constan, D. & Keegstra, K. (2001) *Plant Physiol.* **125,** 1567–1576.
4. Neuhaus, H. E. & Wagner, R. (2000) *Biochim. Biophys. Acta* **1465,** 307–323.
5. Shingles, R., North, M. & McCarty, R. E. (2002) *Plant Physiol.* **128,** 1022–1030.
6. Rolland, N., Droux, M. & Douce, R. (1992) *Plant Physiol.* **98,** 927–935.
7. Ravanel, S., Gakiere, B., Job, D. & Douce, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7805–7812.
8. Flügge, U. I., Fischer, K., Gross, A., Sebald, W., Lottspeich, F. & Eckerskorn, C. (1989) *EMBO J.* **8,** 39–46.
9. The Arabidopsis Genome Initiative (2000) *Nature (London)* **408,** 796–815.
10. Ward, J. M. (2001) *Bioinformatics* **17,** 560–563.
11. Douce, R. & Joyard, J. (1982) in *Methods in Chloroplast Molecular Biology*, eds. Edelman, M., Hallick, R. & Chua, N.-H. (Elsevier Science, Amsterdam), pp. 239–256.
12. Block, M. A., Dorne, A. J., Joyard, J. & Douce, R. (1983) *J. Biol. Chem.* **258,** 13273–13280.
13. Seigneurin-Berny, D., Rolland, N., Garin, J. & Joyard, J. (1999) *Plant. J.* **19,** 217–228.
14. Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J. & Joyard, J. (2000) *Electrophoresis* **21,** 3517–3526.
15. Bradford, M. (1976) *Anal. Biochem.* **72,** 248–254.
16. Chua, N. H. (1980) *Methods Enzymol.* **69,** 434–436.
17. Rolland, N., Droux, M., Lebrun, M. & Douce, R. (1993) *Arch. Biochem. Biophys.* **300,** 213–222.
18. Emanuelsson, O., Nielsen, H. & von Heijne, G. (1999) *Protein Sci.* **8,** 978–984.
19. Tusnády, G. E. & Simon, I. (2001) *Bioinformatics* **17,** 849–850.
20. Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. & Sheen, J. (1996) *Curr. Biol.* **6,** 325–330.
21. Schmidt, H., Dresselhaus, T., Buck, F. & Heinz, E. (1994) *Plant Mol. Biol.* **26,** 631–642.
22. Bölter, B., Soll, J., Hil, K., Hemmler, R. & Wagner, R. (1999) *EMBO J.* **18,** 5505–5516.
23. Miège, C., Maréchal, E., Shimojima, M., Awai, K., Block, M. A., Ohta, H., Takamiya, K., Douce, R. & Joyard, J. (1999) *Eur. J. Biochem.* **265,** 990–1001.
24. Daram, P., Brunner, S., Rausch, C., Steiner, C., Amrhein, N. & Bucher, M. (1999) *Plant Cell* **11,** 2153–2166.
25. Raghothama, K. (2000) *Curr. Opin. Plant Biol.* **3,** 182–187.
26. Walker, D. A. (1976) in *The Intact Chloroplast*, ed. Barber, J. (Elsevier Science, Amsterdam), pp. 235–278.
27. Joyard, J. & Douce, R. (1975) *FEBS Lett.* **51,** 335–340.
28. Kammerer, B., Fischer, K., Hilpert, B., Schubert, S., Gutensohn, M., Weber, A. & Flugge, U. I. (1998) *Plant Cell.* **10,** 105–117.
29. Schäffer, G., Heber, U. & Held, H. W. (1977) *Plant Physiol.* **60,** 286–289.
30. Weber, A., Servaites, J. C., Geiger, D. R., Kofler, H., Hille, D., Groner, F., Hebbeker, U. & Flugge, U. I. (2000) *Plant Cell* **12,** 787–802.
31. Theodoulou, F. L. (2000) *Biochim. Biophys. Acta* **1465,** 79–103.
32. Hortensteiner, S., Vogt, E., Hagenbuch, B., Meier, P. J., Amrhein, N. & Martinoia, E. (1993) *J. Biol. Chem.* **268,** 18446–18449.
33. Leustek, T. & Saito, K. (1999) *Plant Physiol.* **120,** 637–644.
34. Mourioux, G. & Douce, R. (1979) *Biochimie* **61,** 1283–1292.
35. Ortiz-Lopez, A., Chang, H. C. & Bush, D. R. (2001) *Biochim. Biophys. Acta* **1465,** 275–280.
36. Ravanel, S., Cherest, H., Jabrin, S., Grunwald, D., Surdin-Kerjan, Y., Douce, R. & Rebeille, F. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 15360–15365.