The Chloroplast Grana Proteome Defined by Intact Mass Measurements from Liquid Chromatography Mass Spectrometry*

Stephen M. Gómez‡§, John N. Nishio‡, Kym F. Faull§, and Julian P. Whitelegge§¶

Proteomics seeks to address the entire complement of protein gene products of an organism, but experimental analysis of such complex mixtures is biased against low abundance and membrane proteins. Electrospray-ionization mass spectrometry coupled with reverse-phase chromatography was used to separate and catalogue all detectable proteins in samples of photosystem II-enriched thylakoid membrane subdomains (grana) from pea and spinach. Around 90 intact mass tags were detected corresponding to approximately 40 gene products with variable post-translational covalent modifications. Provisional identity of 30 of these gene products was proposed based upon coincidence of measured mass with that calculated from genomic sequence. Analysis of isolated photosystem II complexes allowed detection and resolution of a minor population of D1 (PsbA) that was apparently palmitoylated and not detected in less purified preparations. Based upon observed +80-Da adducts, D1, D2 (PsbD), CP43 (PsbC), two Lhcbs, and PsbH were confirmed to be phosphorylated, and a new phosphoprotein was proposed to be the product of *psbT*. The appearance of a second +80-Da adduct on PsbH provides direct evidence for a second phosphorylation site on PsbH, complicating interpretation of its role in regulation of thylakoid membrane organization and function, including light-state transitions. Adducts of +32 Da, presumably arising from oxidative modification during illumination, were associated with more highly phosphorylated forms of PsbH implying a relationship between the two phenomena. Intact mass proteomics of organellar subfractions and more highly purified protein complexes provides increasingly detailed insights into functional genomics of photosynthetic membranes. ***Molecular & Cellular Proteomics 1:46–59, 2002.***

With many genomes completed and many more in the pipeline it is clear that the post-genomic era has arrived. Considerable attention is now being directed toward defining the function of individual gene products and the inter-relationships between them (functional genomics). Proteomics seeks to catalogue the full complement of the gene products of a cell and the effect of development, environment, and disease upon their expression. Mass spectrometry is driving proteomics, most commonly as a tool to identify proteins separated and visualized on two-dimensional gels. However, such strategies are insensitive to low abundance proteins (1, 2), proteins that are not fully represented on two-dimensional gels (for example, some classes of intrinsic membrane proteins) and to subtle changes in covalent modifications that do not appreciably alter isoelectric point or electrophoretic mobility. To address some of these shortcomings, intact mass proteomics has been proposed (3–8).

The ideal analysis of any protein includes a mass spectrum of the intact molecule to define the native covalent state and its heterogeneity (3, 4). A versatile procedure has been developed for effective electrospray-ionization mass spectrometry (MS)[1] of intact intrinsic membrane proteins purified by reverse-phase chromatography in aqueous formic acid/isopropanol (3). An advantage of this technique is the ability to accurately measure proteins greater than 30 kDa, thereby allowing analysis of the majority of the gene products from any genome. The mass measurements of spinach PS II D1 (38,022 Da; 5 membrane-spanning α-helices (MSH)), D2 (39,419 Da; 5 MSH), and *Halobacterium halobium* bacteriorhodopsin holoprotein (27,052 Da; 7 MSH) were within 0.01% of calculated theoretical values, setting a benchmark standard for analysis of intrinsic membrane proteins (3). The high mass accuracy of this technique for larger hydrophobic proteins has been demonstrated by analysis of a His_6-tagged *Escherichia coli* lactose permease (*lacY*; 47,357 Da; 12 MSH) (5, 7), His_6-tagged *Vibrio parahaemolyticus* Na^+/galactose cotransporters (*sglT*; 60,676 and 90,544 Da; 14/15 MSH) (9), and the α-subunit of the rat Na^+/K^+-ATPase (*atn2*; 112,344 Da; 10 MSH).[2]

Strict translation of a published gene sequence is usually not sufficient to match a mass to a gene, and genome sequence

From §The Pasarow Mass Spectrometry Laboratory, Department of Psychiatry and Biobehavioral Sciences, The Neuropsychiatric Institute and the Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90095 and the ‡Department of Botany, University of Wyoming, Laramie, Wyoming 82071

Received, August 21, 2001, and in revised form, October 11, 2001
Published, MCP Papers in Press, October 12, 2001, DOI 10.1074/mcp.M100007-MCP200

[1] The abbreviations used are: MS, mass spectrometry; MS-MS, tandem mass spectrometry; HPLC, high-performance liquid chromatography; PS, photosystem; MSH, membrane-spanning α-helices; LC-MS, liquid chromatography-mass spectrometry; IMT, intact mass tag.

[2] J. P. Whitelegge and S. J. Karlish, unpublished data.

manipulation is required before agreement with measured masses is achieved. Varietal differences, DNA sequencing errors, post-transcriptional and post-translational modifications, as well as protein damage, must all be considered. Spectra frequently reveal lesser quantities of other molecular species that can usually be equated with covalently modified subpopulations of the dominant proteins (3, 4).

Toward a complete description of the thylakoid proteome, the closely appressed membranes of the granal stacks were prepared using their known resistance to solubilization by Triton X-100, resulting in a subfraction highly enriched in the polypeptides of PS II. Using this simplified starting material it was possible, using LC-MS, to record the masses of all detectable polypeptides in PS II-enriched membrane preparations from spinach and pea. Proteins were identified provisionally based upon coincidence of measured mass with that calculated from sequences in the data base and their predicted elution from the HPLC column based upon calculated hydrophobicity. The heterogeneity of the larger PS II subunits was revealed illustrating the accuracy and resolution afforded by electrospray-ionization-MS. The observation of light-dependent double phosphorylation of PsbH demonstrates potential advantages and pitfalls of the approach for examining relative expression and steady-state native modifications. Analysis of subfractions of the proteome using this technique decreased the genomic coverage but allowed resolution of several, otherwise unrecognized, native modifications that may be functionally significant.

EXPERIMENTAL PROCEDURES

Leaves from 3-week-old greenhouse-grown pea (*Pisum sativum* cv. Alaska) plants, 6-month-old tobacco (*Nicotiana tabacum* cv. Samsung) plants, and spinach obtained from local market sources were used for preparation of PS II-enriched membranes (10). Samples (80 μg of chlorophyll or 250 μg of protein; pea, $n = 3$; spinach, $n = 7$; tobacco $n = 2$) were prepared by acetone precipitation prior to dissolution in 60% HCOOH (diluted from 90% ACS grade; Fisher). Primary reverse-phase chromatography was performed as described previously (3). The poly(styrene-divinylbenzene) copolymer (Polymer Labs PLRP/S; 5 μm $\times$ 300 Å; 2.1 $\times$ 150 mm) stationary-phase column was eluted at a flow rate of 100 μl/min at 40 °C. The primary gradient (Buffer A, 0.1% trifluoroacetic acid/water; Buffer B, 0.1% trifluoroacetic acid/acetonitrile) eluted extrinsic polypeptides and predominantly small to moderately sized intrinsic proteins. The column was equilibrated in 5% Buffer B followed by a stepped linear gradient from 5 to 25% Buffer B between 5 and 10 min after injection, 25 to 75% Buffer B between 10 and 130 min, and 75 to 100% Buffer B between 130 and 150 min.

The spinach D1 spectra (see Fig. 2*A*) was obtained by acetone precipitation of PS II reaction centers (100 μg of protein, prepared according to Ref. 11) that were dissolved in 60% formic acid and subjected to reverse-phase chromatography with the gradient described above, except using Buffer A and Buffer C (0.05% trifluoroacetic acid in 1:1 acetonitrile/2-propanol, v/v). The addition of 2-propanol improves the elution efficiency for larger intrinsic membrane proteins.

A secondary gradient (Buffer D, 60% HCOOH; Buffer E, 2-propanol) was used to elute the very hydrophobic PS II polypeptides that remained bound to the column after the primary chromatographic elution in the AB buffer system described above (3, 11). The column was equilibrated in 95% Buffer D, 5% Buffer E, prior to linear gradient elution to 100% Buffer E over 55 min. Six primary runs were used to accumulate material for each secondary elution.

Mass spectra were recorded on a PerkinElmer Life Sciences Sciex API III+ triple-quadrupole mass spectrometer with an Ionspray™ source (Applied Biosystems, Foster City, CA) as described (3). The instrument was scanned from 600–2300 *m/z* with a step size of 0.3 *m/z* and a dwell of 1 ms giving a total scan time of 6 s. An orifice potential of 65 V was used in all experiments. The computations of measured protein molecular mass were made using MacSpec 3.3 and zero-charge molecular mass reconstructions using BioMultiView 1.3.1 software (Applied Biosystems, Foster City, CA). Calculated average molecular masses were generated from translated published gene sequences (GenBank™) or published protein sequences (PIR or Swiss-Prot) using PeptideMass (expasy.cbr.nrc.ca/tools/peptide-mass.html). Post-translational modifications of thylakoid proteins were predicted by comparison of published modifications of orthologs.

Predictions of the HPLC retention time of proteins under 40 kDa were made by searching the ARATH (*Arabidopsis thaliana*) subset of the Swiss-Prot and TrEMBL data bases with TagIdent (ca.expasy.org/tools/tagident.html). The ProtParam tool at ExPASy was used to calculate a predicted hydrophobicity (GRAVY; grand average of hydropathicity) of each of the proteins found in the TagIdent search. The GRAVY results from *Arabidopsis thaliana* were plotted against the predicted masses and compared with the observed pea HPLC retention times. The results of the TagIdent search (performed in July 2001) represent a crude data set because of the lack of, or incorrect, annotation of the *A. thaliana* genomes, especially of nuclear-encoded chloroplast proteins that are proteolytically trimmed of a substantial N-terminal leader sequence prior to thylakoid insertion *in vivo*.

RESULTS

Thylakoid membrane proteins were stripped of lipids, chlorophyll, and other pigments by precipitation with acetone and dissolved in 60% formic acid for the primary reverse-phase HPLC analysis coupled with electrospray-ionization mass spectrometry (LC-MS; see Ref. 3). Many polypeptides elute efficiently under standard conditions of 0.1% trifluoroacetic acid with increasing acetonitrile concentration. Averaging the elution profiles of several experiments allows a generalized elution map to be generated (Fig. 1). The trace shown follows total ion production across the scanned mass range such that peaks appear as molecules of different masses elute, rather like a protein UV elution profile except that relative abundance can be biased by the potential for different proteins to have different ionization efficiencies. Generally, the abundance of smaller proteins is exaggerated, because they ionize more efficiently than larger ones.

The very hydrophobic proteins elute from the column with lowered efficiency such that a substantial proportion remain bound to the column. Thus more hydrophobic PS II core polypeptides, including larger ones such as D2, CP43, and CP47, and small ones such as PsbM, are variably detected in the primary elution profile for appressed membrane fractions. The addition of isopropanol to acetonitrile (Buffer C) improved elution efficiency, though optimal spectra of D1 and D2 were obtained only by loading larger quantities of highly purified PS II reaction-center preparations (a subfraction of the appressed thylakoids used for the primary elution described above) to

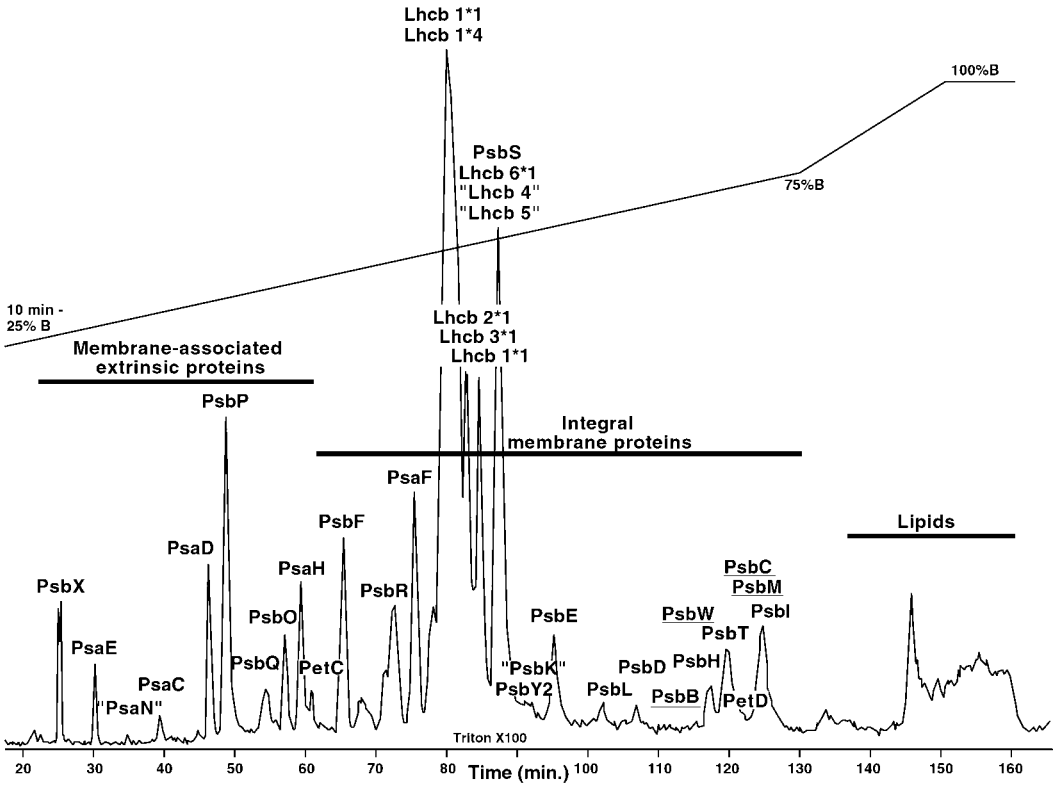

FIG. 1. **Generalized reverse-phase HPLC total ion chromatogram elution profile of PS II-enriched thylakoid membranes.** Polypeptides were extracted from appressed membrane-enriched subfractions of pea thylakoid membranes. Peak height is dependent upon ionization efficiency of individual proteins and efficiency of elution from the reverse-phase column. The column was equilibrated in 0.1% trifluoroacetic acid, 5% acetonitrile; acetone-precipitated proteins were dissolved in 60% formic acid for loading, and the column was eluted with a gradient as shown (B = 0.1% trifluoroacetic acid in acetonitrile). Eluent is directed to the ionization source for on-line mass spectrometry, a spectrum being recorded every 6 s. The elution positions of identified proteins are labeled. *Underlined* labels indicate inferred elution positions from other experiments. *Quotations* indicate no published pea or spinach DNA sequence, and identification is based upon similar mass and predicted retention time of orthologs in other species. Elution positions of residual Triton X-100 and lipids are also labeled.

reduce the total number of proteins in the sample (Fig. 2*A*). Efficient elution of the more hydrophobic PS II polypeptides was achieved in 60% formic acid with elution by isopropanol (Buffers D and E), but chromatographic resolution is impaired, and less complex mixtures are preferable (3). The generalized primary elution profile for appressed membranes (Fig. 1) was annotated with peak identities assigned by comparison of measured intact masses to values calculated based upon gene sequences and available information on post-transcriptional and post-translational modifications. There are few proteins for which all this information is available. However, specific genes in related species (orthologs) are generally highly conserved in sequence, and known modifications in one ortholog were useful for predicting the same modification in another.

Extrinsic polypeptides, including those of the oxygen-evolving complex, elute early whereas more hydrophobic intrinsic polypeptides elute later in the primary elution profile. The masses of detected proteins and their proposed identities are presented in Table I in the order of their elution. Each significant mass spectrometric signal yielding a molecular weight is considered an intact mass tag (IMT) and compared with the data base. Some IMTs are easily assigned, because they correspond closely to well studied proteins whose mass can be calculated with great confidence based upon previous characterization. The native masses of some of the proteins that have been previously identified match very well with the IMT and are so indicated in Table I. Others are assigned based upon the pure coincidence of measured and calculated masses, with confidence being modulated by the number of other proteins of similar mass and similar hydrophobicity within the sample. In special cases it was necessary to compare all known orthologs of a particular protein and use phylogenetic conservation of sequence to identify probable sequence errors.[3] The calculated masses presented here are based upon sequences from data base entries without such sequence correction. Of the 90 IMTs recognized in the primary/secondary elution profiles (see Tables I and II), 40 genes were assigned, leaving 30 IMTs unassigned. Some of these likely correspond to the products of genes whose function

[3] J. P. Whitelegge, S. M. Gómez, and K. F. Faull, manuscript in preparation.

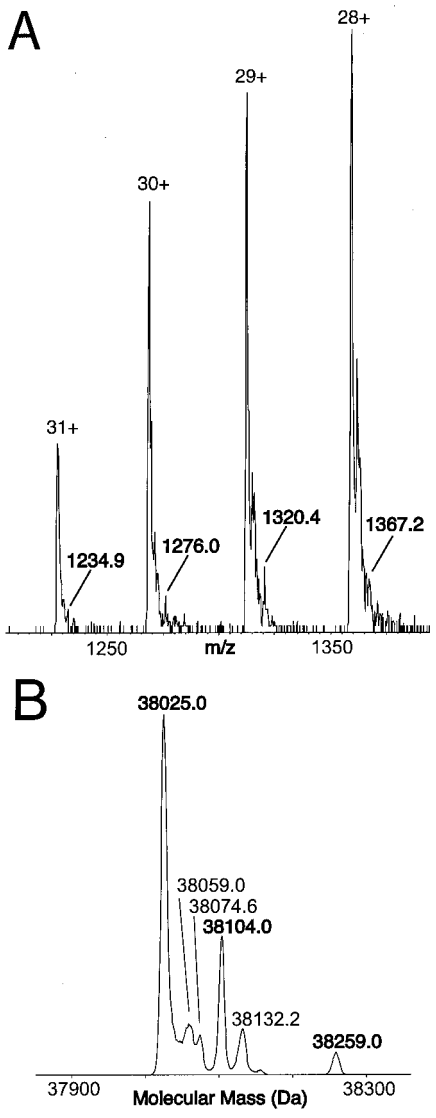

FIG. 2. **Electrospray-ionization mass spectrometry of minor thylakoid membrane protein isoforms.** *A*, mass spectrum of spinach D1 protein. A segment of the mass spectrum showing four multiply charged ions derived from D1 is shown to illustrate signal to noise and the detection of a minor palmitoylated species. Each ion is labeled with its charge state and the measured *m/z* of the minor species. The molecular weight of the uncharged molecule is calculated by multiplying *m/z* by the charge state and subtracting the mass of the charges (protons). Measured molecular weight is averaged from all detectable signals across the complete mass spectrum (Hypermass, PE Sciex; Applied Biosystems, Foster City, CA). *B*, zero-charge molecular weight spectrum of spinach D1 was reconstructed from corresponding mass spectrum by computation (BioMultiView, PE Sciex; Applied Biosystems, Foster City, CA).

remains as yet unidentified or proteins for which there is no gene sequence data in pea or spinach.

The majority of proteins found in the preparations were the membrane-bound PS II proteins, but extrinsic PS I and PS II proteins were also observed. The abundance of extrinsic proteins was dependent upon whether the appressed membranes were NaBr-washed after Triton treatment. Detectable quantities of some of the extrinsic components of PS I were measured even in washed preparations. Varying proportions of the more hydrophobic membrane-bound PS II proteins remain immobilized on the column under the primary elution gradient conditions used to generate the data in Table I but can be eluted efficiently with a second gradient. Table II shows the IMTs obtained from experiments using the formic acid-isopropanol system (solvents D and E) for the secondary elution profile applied to pea appressed-membrane preparations. Although several intrinsic membrane-bound proteins observed in the primary elution profile are also evident in the secondary elution, no extrinsic PS I or PS II proteins were observed in the secondary elution.

Some IMTs were observed only when larger quantities of more highly purified preparations were analyzed. A segment of the spinach D1 (PsbA) ion series (Fig. 2*A*) with 31–28 charges is shown to illustrate the signal to noise ratio in the mass spectrum for the major and minor isoforms. Molecular weight is obtained by multiplying the observed *m/z* by the charge state and subtracting the mass of the charging species (number of protons), a process that is automated for presentation of averaged molecular weight spectra (Fig. 2*B*). Significantly, the minor signals (*m/z* 1234.9, 1276.0, 1320.4, and 1367.2; see Fig. 2*A*) in the mass spectrum, corresponding to the palmitoylated isoform of D1 described previously (12), were not discernable when the more complex appressed membrane grana subfractions were analyzed. Thus by working with more highly purified complexes we detect otherwise unseen minor components, of potential functional significance. The predominant mass peak in the reconstructed molecular weight spectrum of spinach D1 is in good agreement with the sequence translated from the gene, with removal of fMet and acetylation of Thr^2 (13), as well as C-terminal processing (14) (38025.0 Da observed; 38020.6 Da calculated; difference between observed and calculated masses, Δ = 0.012%). The molecular weight reconstruction also shows the phosphorylated species (38104.0; 38100.6; 0.009%) and a palmitoylated species (38259.0; 38254.6; 0.012%), as well as other signals (38059.0; 38132.2) that probably arise from protein formylation during sample preparation. The data obtained with the acetonitrile/isopropanol system (solvents A and C) provides improved overall isoform resolution over the originally presented spectrum of spinach D1 obtained with the formic acid/isopropanol system (solvents D and E, compare spectrum to Ref. 3).

Based upon results from the secondary gradient elution of appressed thylakoids from pea, electrospray-ionization spectra are reported for the four large core subunits of PS II. Pea D1 (PsbA) is apparently processed as in spinach with N- and C-terminal trimming and *N*-acetylation (Fig. 3*A*, 38033.0 Da measured; 38033.5 Da calculated; Δ = 0.002%) with a lesser subpopulation apparently phosphorylated (38118.0; 38113.5; 0.012%), though no evidence was observed for the palmitoy-

TABLE I
Generalized primary reverse-phase elution map of intact mass tags from appressed thylakoid membrane subfractions

IMT[a] pea (± S.D.) (n = 4 experiments)	Calculated masses[b] (Da)	Δ[c]	IMT[a] spinach (± S.D.) (n = 7 experiments)	Calculated masses[b] (Da)	Δ[c]	ID[d]
			3296.5 ± 0.2	3297.9	Z	PsbX + O
3187.3 ± 0.2			3279.7 ± 0.4	3281.9	Z	PsbX
3485.2 ± 0.4			3163.2 ± 0.3			
			3000.0 ± 0.7			
2088.7 ± 0.2						
9894.3 ± 1.2			9728.9 ± 0.9	9729.0	X	PsaE
9824.4 ± 1.1			9657.7 ± 2.0	9657.9	X	PsaEΔAla
4028.8 ± 0.5						
9721.3 ± 0.9						"PsaN"
			9150.5 ± 1.3			
			8641.0 ± 1.6			
8848.4 ± 0.9	8849.3	X				PsaC
17936.3 ± 2.3			17871.8 ± 1.4	17872.6	X	PsaD + O
17921.3 ± 1.4			17856.7 ± 2.1	17856.6	X	PsaD
			20225.4 ± 1.5	20225.6	X	PsbP + O
20265.0 ± 1.6	20264.6	X	20209.6 ± 2.0	20209.6	X	PsbP
16375.2 ± 2.0			16522.7 ± 2.9	16521.8	X	PsbQ
			17829.4 ± 2.1			a + 2F
			17801.7 ± 2.7			a + F
			17774.6 ± 2.3			a
16143.4 ± 2.6						b + Ac
16102.0 ± 1.0						b
12361.4 ± 1.3						
26525.0 ± 1.8	26525.7	X	26532.1 ± 2.9	26661.8	1	PsbO
10276.3 ± 0.9			10381.6 ± 1.2	10381.8	X	PsaH
15564.7 ± 2.0						
12941.4 ± 2.5						
			13815.4 ± 1.8			
			22037.2 ± 2.0			
			18956.5 ± 3.2	18954.6	X	PetC + O
19074.7 ± 2.4	19074.8	X	18940.4 ± 2.3	18938.6	X	PetC
11320.3 ± 1.3			10764.8 ± 0.5			
29474.1 ± 2.6						
			8596.3 ± 0.8			
6927.9 ± 1.3			6924.0 ± 1.1			c + Ac + O
6911.6 ± 0.7			6907.3 ± 0.9			c + Ac
6869.2 ± 0.2			6869.0 ± 1.5			c
4411.1 ± 1.2	4411.2	X	4424.7 ± 0.5	4425.2	X	PsbF + O
4394.9 ± 0.3	4395.2	X	4409.2 ± 0.6	4409.2	X	PsbF
15178.6 ± 1.0			15171.8 ± 2.4			
22230.6 ± 1.8						
21781.0 ± 1.8						d + Ac
			21677.3 ± 1.6			d + O
21739.0 ± 1.3			21661.2 ± 3.6			d
21441.8 ± 2.8						
			4854.8 ± 0.7			
25285.3 ± 3.3			25298.6 ± 2.9			
10205.7 ± 1.4			10234.9 ± 1.4	10235.7	X	*PsbR*
8389.0 ± 0.4			9685.3 ± 0.9			
21782.7 ± 2.8			22345.6 ± 2.5			
17255.3 ± 1.2			17272.3 ± 3.0	17275.1	Y	PsaF
3983.3 ± 0.3						
4585.6 ± 0.6						
24574.8 ± 1.3			24839.7 ± 2.8			
24972.0 ± 2.2	24972.3	X	24760.2 ± 2.3			Ac-Lhcb1*Ps4
24981.1 ± 1.9			25019.1 ± 3.4	25020.4	X	Ac-Lhcb1*So1
25036.4 ± 2.2	25036.3	X				Ac-Lhcb1*Ps2 + P
24959.1 ± 1.5	24956.3	Y	24946.3 ± 2.9			Ac-Lhcb1*Ps2
24836.4 ± 2.1	24837.1	X				Ac-Lhcb2*Ps1

TABLE I—*continued*

IMT[a] pea (± S.D.) (*n* = 4 experiments)	Calculated masses[b] (Da)	Δ[c]	MT[a] spinach (± S.D.) (*n* = 7 experiments)	Calculated Masses[b] (Da)	Δ[c]	ID[d]
24750.7 ± 1.9			24936.2 ± 2.6			
24839.1 ± 1.4			25083.7 ± 3.0			
24774.2 ± 1.9			25005.2 ± 3.2			
22913.4 ± 1.6			22856.1 ± 3.4			
22842.5 ± 1.6			22814.0 ± 1.9	22813.0	X	Lhcb6*So1
22302.9 ± 1.8						
22233.8 ± 1.6						
24332.7 ± 2.3	24330.8	X	24323.3 ± 2.3			Lhcb3*Ps1
22952.9 ± 3.1						
24345.9 ± 2.0						
28690.8 ± 2.3			28071.8 ± 3.9			"Lhcb4"
26262.5 ± 2.4						
26562.8 ± 1.9			27067.4 ± 2.7			"Lhcb5"
22169.8 ± 1.8			22568.6 ± 2.1	22568.3	X	Ac-PsbS
4540.3 ± 0.4			4960.4 ± 0.5	4960.8	X	PsbY2
4482.0 ± 0.0						"PsbK"
23345.8 ± 2.0			23324.0 ± 2.9			
27379.2 ± 2.6			23279.4 ± 2.9			
9283.5 ± 0.9	9283.4	X	9255.4 ± 0.9	9255.4	X	*PsbE*
17934.6 ± 2.2			18056.5 ± 1.8			e + O
17919.6 ± 1.1			18040.2 ± 1.6			e
4409.3 ± 0.5			3972.1 ± 0.4			
4272.7 ± 0.6			3866.0 ± 0.2			
4367.1 ± 0.1	4366.0	Y	4365.9 ± 0.8	4366.0	X	*PsbL*
4759.7 ± 0.2						
39481.7 ± 2.8	39475.9	Y				Ac-PsbD
39439.6 ± 2.5	39437.5	X				PsbD
			56032.2 ± 6.7	56064.0		Ac-PsbB
			10101.1 ± 1.6			
			7774.0 ± 0.6	7774.9	Y	PsbH + 2P + O
7792.4 ± 0.5	7823.0	2	7694.3 ± 0.6	7694.9	X	PsbH + P + O
7713.9 ± 0.8	7743.0	2	7613.9 ± 1.1	7614.9	Y	PsbH + O
7857.4 ± 1.1	7887.0	2	7758.8 ± 0.6	7758.9	X	PsbH + 2P
7777.3 ± 0.9	7807.0	2	7679.3 ± 0.9	7678.9	X	*PsbH + P*
7697.3 ± 1.0	7727.0	2	7599.5 ± 1.2	7598.9	X	*PsbH*
5913.2 ± 0.4			5927.7 ± 0.4	5927.7	X	PsbW
			17329.3 ± 1.6	17329.6	X	PetD + O
17359.1 ± 1.2	15136.0	3	17313.7 ± 2.0	17313.6	X	PetD
4140.6 ± 0.6	4112.5	4				fM-PsbT + P
4060.8 ± 0.7	4032.5	4	3849.0 ± 0.4	3849.7	Y	*fM-PsbT*
4284.9 ± 0.3						
50205.9 ± 8.3	50204.9	X				Ac-PsbC
			3809.8 ± 0.6	3810.6	Y	*fM-PsbM*
			4212.0 ± 0.6	4211.9	X	fM-PsbI + O
4210.3 ± 0.6	4209.8	Y	4195.8 ± 0.5	4195.9	X	*fM-PsbI*

[a] IMTs assembled in order of elution during the primary reverse-phase chromatographic elution. The mean ± S.D. of *n* experiments is presented.

[b] Calculated average mass of the uncharged assigned gene product shown in ID column.

[c] Δ, % difference between expected and observed masses; X, Δ < 0.01%; Y, Δ = 0.01–0.02%; Z, Δ = 0.02–0.07%. 1, IMT 26532.1 is in the mass range and retention time for PsbO, except the mass does not match the published sequence. The mass difference of 129 Da could reasonably be expected to be because of one or more amino acid changes. Removal of a single Q from the C terminus (26533.7 calculated) results in excellent agreement, for example. 2, the pea PsbH sequence has at least two amino acid changes from the published sequence. We have identified one (A18V) by MS/MS (not shown). 3, the pea *petD* sequence does not include the first exon and erroneously starts with an AUG codon in the second exon. A chimeric protein translated from the first exon of the spinach sequence and the second exon from pea gives a calculated mass of 17369.7 Da. The mass difference is consistent with changing either Pro^{8} or Pro^{12} encoded in exon 1 of spinach *petD* to Ser. 4, the published pea *psbT* gene sequence appears to have an in-frame deletion that is currently being confirmed.

[d] Protein IDs in quotes indicate assignments made by retention times consistent with membrane localization and similar masses from published sequences from other plants. IDs in italic type indicate that a tobacco IMT (not shown) was correlated with a published tobacco gene sequence and that the protein eluted at a similar time to the pea and spinach orthologs. O, oxygen; P, phosphate; Ac, acetyl; F, formyl; fM, *N*-formylmethionine. Five unassigned post-translationally modified masses are labeled in lower case letters (a–e) in order of elution. All accession numbers are from GenBank™, unless otherwise noted. Pea: *psaC*, X13157; *psbP*, X15552; *psbO*, X15350; *petC*, X63065; *psbF*, *psbJ*, *psbE*, *psbL*, X15767; *Ihcb1*4*, X56338; *Ihcb1*2*, K02067; *Ihcb2*1*, X57082; *Ihcb3*1*, X69215; *psbD*, M27309; *psbB*, *psbH*, *psbT*, AF153442; *petD*, X00535; *psbA*, M11005; *psbC*, M27309; *psbM*, D12535; PsbI (see Ref. 39). Spinach: Psb*X*, pir# SO3277 (unidentified *X* changed to Cys); *psaE*, X14018; *psaD*, X14617; *psbP*, X05511; *psbQ*, X05512; *psbO*, X05548; *psaH*, X16858; *petC*, X06244; *psbF*, *psbE*, *psbL*, M35673; *psbR*, J03887; *psaF*, X13133; *Ihcb1*1*, X14341; *Ihcb6*1*, Z25886; *psbS*, X68552; *psbY2*, AF060198; *psbB*, NP054960; *psbH*, NP054963 amino acids 8–79 (earlier entries initiate from the second methionine 7, but internal sequence errors are corrected in this sequence); *psbT*, P37259; *psbM*, NP054926; *psbI*, NP054916; *psbW*, X85038; *petD*, X07106.

TABLE II

Secondary reverse-phase elution map of intact mass tags from appressed-membrane subfractions of pea thylakoids (n = 2)

IMT[a] mass (± range)	Calculated masses[b]	Δ[c]	ID[d]	Modifications[e]
20261.3				
22233.6	*22233.8		Lhca?	
22841.1	*22842.5		Lhcb6	NH_2-$A^{??}$
23345.0 ± 0.2	*23345.8		Lhca?	
24346.9 ± 0.5	24330.8	1	Lhcb3	NH_2-G^{43}
24751.2 ± 0.5	*24750.7		Lhcb1/2	Ac-$R^{??}$
24838.3 ± 0.1	24837.1	X	Lhcb2	Ac-R^{38}
24958.6	24956.3	X	Lhcb1*2	Ac-R^{38}
24972.2	24972.3	X	Lhcb1*4	Ac-R^{28}
24981.2 ± 0.2	*24981.1		Lhcb1/2	Ac-$R^{??}$
25036.1	25036.3	X	Lhcb1*2	Ac-R^{38}, HPO_3-T^{40}
25061.4	$ 25061.1		Lhcb1/2	Ac-$R^{??}$, HPO_3-$T^{??}$
26563.5 ± 0.5	*26562.8		"Lhcb5"	NH_2-?
26526.9				
27379.2	*27379.2			
28691.7 ± 0.2	*28690.8		"Lhcb4"	Ac-?
9283.3	9283.4	X	PsbE	
55895.1 ± 0.2	55902.0	Y	PsbB	Ac-G^2
17919.2 ± 0.1			a	
221678.6 ± 1.6	*22169.8	X	PsbS	Ac
38031.4 ± 1.6	38033.6	X	PsbA.1[f]	Ac-T^2, A^{344}-COOH
38111.1	38113.6	X	PsbA.1[f]	Ac-T^2, HPO_3-T^2, A^{344}-COOH
38040.7	38040.9	X	PsbA.2[f]	Ac-T^2, A^{344}-COOH
38121.3	38120.9	X	PsbA.2[f]	Ac-T^2, HPO_3-T^2, A^{344}-COOH
50201.4 ± 5.6	50204.9	X	PsbC	Ac-T^{15}
50285.0	50284.9	X	PsbC	Ac-T^{15}, HPO_3-T^{15}
4366.2 ± 1.6	4366.0	X	PsbL	NH_2-T^2
39444.2 ± 3.2	39437.5	X	PsbD	
39487.0	39479.5	Y	PsbD	Ac-T^2
39520.7	39517.5	X	PsbD	
39550.0	39559.5	Z	PsbD	Ac-T^2, HPO_3-T^2
4026.0 ± 1.6	4026.7	X	PsbJ	Ac-A^2
4157.3			f	
4174.3			f	+O
4060.5 ± 0.3	4032.5	2	PsbT	fM
4141.5 ± 0.1	4112.5	2	PsbT	fM, + HPO_3
4209.8 ± 0.1	4209.8	X	PsbI	fM
3782.9	3783.6	Y	PsbM	fM
6641.5 ± 0.1	6582.8	3	Ycf9	fM

[a] IMTs assembled in order of elution during the secondary reverse-phase chromatographic elution. Grouped IMTs indicate co-elution. IMTs without a range were observed in only one or the other experiment.

[b] Calculated average mass of the uncharged assigned gene product shown in ID column (Da). *, mass observed in Table I but with no published gene sequence from which to predict a mass. $, modification of a (*) mass observed in Table I.

[c] Δ, % difference between expected and observed masses; X, Δ < 0.01%; Y, Δ = 0.01–0.02%; Z, Δ = 0.02–0.07%. 1, the mass and relative elution time of IMT 24347.4 Da is consistent with Lhcb3. The difference may be because of one of several possible single amino acid substitutions. 2, the published pea *PsbT* gene sequence appears to have an in-frame deletion (see text). 3, the IMT 6641.5 Da elutes at a time consistent with that predicted for Ycf9 (a conserved protein in plant chloroplast genomes of unknown function) and has similar mass. The discrepancy may be because of published errors in the DNA sequence or because of allelic differences in the cultivars used (Alaska, this study; Rosakrona, M27309).

[d] Protein IDs in quotes indicate assignments made by retention times consistent with membrane localization and similar masses from published sequences from other plants.

[e] O, oxygen; Ac, acetyl; fM, *N*-formylmethionine; HPO_3, *O*-phosphorylation. Two unidentified masses that are observed to be modified are labeled in lower case letters (a, f). Superscripts indicate position from Met as amino acid 1. Sequence accession numbers are listed in the footnotes to Table I, except for Ycf9 (M27309).

[f] In this experiment (*Pisum sativum* cv. Alaska) there are potentially two isoforms of D1. The first isoform (38031.4 Da) is within 0.006% of the mass calculated from the published sequence (M11005). The second isoform (38040.7 Da) is within 0.0004% of the mass observed in previous LC-MS experiments (15). Both isoforms are consistent with removal of fMet, acetylation of residue 2, and C-terminal removal of the last nine amino acids, and both have phosphorylated subpopulations. There are several single amino acid substitutions that could account for the 7-Da difference in isoform masses. Further experiments using various pea cultivars are necessary to confirm this observation.

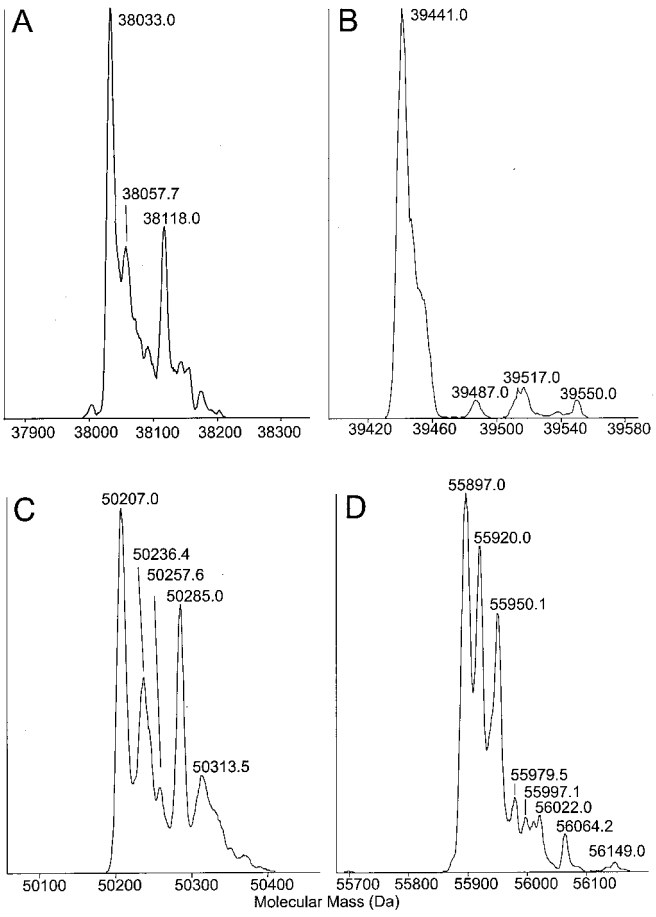

FIG. 3. **Large PS II membrane proteins elute efficiently during secondary elution of reverse-phase column.** Membrane proteins that elute with low efficiency during the primary elution (Buffers A and B; see "Experimental Procedures") accumulate on the column and are eluted with a secondary buffer system (Buffers D and E; see "Experimental Procedures"). This subfractionation technique allowed electrospray-ionization mass spectrometry of the larger PS II subunits from the cruder appressed membrane preparation avoiding the need to prepare reaction-center subfractions, as used in Fig. 2. Note, however, that a palmitoylated isoform of pea D1 was not resolved. Zero-charge molecular weight spectra were prepared as for Fig. 2. *A*, pea D1; *B*, pea D2; *C*, pea CP43; *D*, Pea CP47.

lated isoform after elution with formic acid/isopropanol as we observed in spinach under the same solvent conditions (3). The minor species at 38057.7 may correspond to a singly formylated form of D1 (38061.5 Da calculated; 0.011%). D2 (PsbD) is apparently predominantly non-acetylated after fMet removal (Fig. 3*B*, 39441.0; 39437.5; 0.009%) and has a phosphorylated isoform (39517.0; 39517.5; 0.001%). Two other minor species potentially correspond to the acetylated isoform (39487.0; 39479.5; 0.019%) and the acetylated phosphorylated isoform (39550.0; 39559.5; 0.024%) observed previously in spinach (13). The lower mass coincidence in the latter two species is presumably a consequence of the weaker ion currents because of the low abundance of these isoforms in the sample. The measured masses of pea D1 and D2 reported herein are lower than values reported previously (3, 15), possibly because of allelic differences in the different pea cultivars used (Table II). The other two large PS II core subunits are CP43 (PsbC) (Fig. 3*C*) and CP47 (PsbB) (Fig. 3*D*). Pea CP43 is trimmed, removing 14 amino acids at the N terminus and acetylated (50207.0; 50204.9; 0.004%), and a population is phosphorylated (50285.0; 50284.9; 0.0002%) in agreement with MS-MS data taken from N-terminal tryptic phosphopeptides of spinach CP43 (13). The major pea CP47 species is apparently *N*-acetylated after fMet removal (55897.0; 55902.0; 0.009%). Other heterogeneity may potentially arise from oxidation, formylation, allelic sequence differences, or the existence of multiple copies of the gene in pea or spinach.

Post-translational modifications can provide useful functional information, because these are often regulatory. Scanning the intact mass tags for characteristic mass differences allows the identification of specific modifications. Phosphorylation increases mass by 80 Da, and several tags were indeed separated by this increment. IMTs identified as D1, D2, CP43, two Lhcbs, and PsbH were observed with the +80-Da tag confirming their partial phosphorylation *in vivo*. Furthermore, a tag provisionally assigned as PsbT also exhibited the +80 tag suggesting the presence of a previously unidentified thylakoid phosphoprotein. Moreover, PsbH had a second +80 tag providing convincing evidence to support a second phosphorylation site on spinach and pea PsbH *in vivo*, as matrix-assisted laser desorption/ionization time-of-flight analysis of spinach (16) and tobacco[4] has shown previously. Mass spectrometric sequencing provided evidence for phosphorylation of Thr^2 and Thr^4 on a tryptic peptide recovered from proteolyzed *A. thaliana* thylakoids matching the N-terminal sequence predicted for PsbH (17), consistent with dual phosphorylation of the intact PsbH protein. Molecular weight reconstructions of spinach PsbH are shown in Fig. 4 demonstrating the utility of LC-MS in examining physiological changes to the native covalent state of the protein under different light conditions. Labeling studies provided compelling evidence for phosphorylation of thylakoid polypeptides smaller than PsbH (18). Other common modifications that are observed include covalent oxidation of methionine (+16 Da) and formylation as a result of formic acid treatment (+28 Da), as well as non-covalent adducts such as trifluoroacetic acid (+113 Da). IMTs arising from these adducts were ignored unless considered relevant to the *in vivo* state, *i.e.* reproducible oxidation.

DISCUSSION

Proteomics seeks to reconcile protein data against genomic data to provide functional insights across the entire complement of expressed information. Most proteomics studies have focused upon identification of proteins after their fragmentation into small pieces that are subsequently matched to segments of genes. For example, the proteins of

[4] Unpublished data.

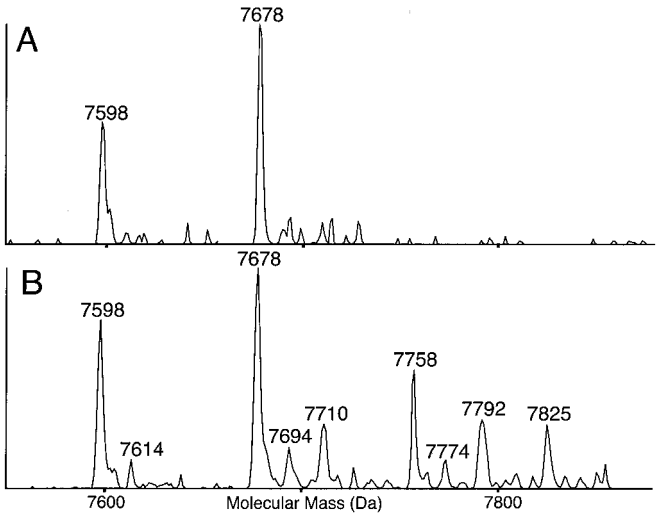

FIG. 4. **Light-dependent modification of phosphoprotein PsbH.** Visibly undamaged leaves from spinach obtained at the market were illuminated prior to preparation of thylakoid membranes. *A*, 50 μmol of photons m^{-2}·sec^{-1} for 9 h. The non- and singly phosphorylated forms of PsbH are labeled (7598 Da, 7678 Da). *B*, 250 μmol of photons m^{-2}·sec^{-1} for 45 min. The non- and singly phosphorylated forms are labeled, as is the doubly phosphorylated form (7758 Da). Single (7614 Da, 7694 Da, 7724 Da) and double (7710 Da, 7792 Da) oxidations are indicated. The peak at 7825 Da appears to carry four oxygens. The appearance of doubly oxidized forms of PsbH correlates with increased phosphorylation.

the thylakoid lumen have been separated by one- or two-dimensional gel electrophoresis, and stained spots were identified by use of peptide-mass tags, sequence tags, and N-terminal Edman sequence data (19–21).

Using an alternative approach, we have compiled a list of intact proteins detected in Triton-resistant thylakoid-membrane subfractions enriched in PS II (10) using LC-MS. The list of IMTs defines the appressed thylakoid proteome, at least for proteins of sufficient abundance to permit detection. Each IMT has been examined carefully by comparison of measured mass to masses calculated from published gene sequences and predicted elution time from the HPLC column based on predicted hydrophobicity. Based upon these comparisons, the identity of many of the IMTs has been proposed thereby allowing consideration of proteome coverage with respect to that predicted based upon genomic information and the body of the current knowledge of protein modification.

Because complete genome information is not available for spinach or pea, the *A. thaliana* genome was examined (23). The predicted total number of *A. thaliana* chloroplast proteins less than 40 kDa is presented in Table III. The TagIdent program identified more than 700 proteins that are predicted to be less than 40 kDa. The chloroplast subset of these proteins contains 133 members, 51 of which are known to be associated with the thylakoid membranes (either intrinsically or extrinsically). The search performed in July 2001 is an underestimate of the number of proteins in *A. thaliana* under 40 kDa because of incomplete annotation of the *A. thaliana* genomes. Several chloroplast proteins encoded in the nucleus (*i.e.* PsaH, PsbT, PsbW, and several light-harvesting proteins) were not returned by the TagIdent searches, because the chloroplast-targeting peptide is incorrectly annotated in a large number of the genes in the genomes. The mass range of several of these incorrectly annotated proteins is greater than 40 kDa, even though the processed protein is substantially less than 40 kDa. No attempt was made to correct the annotations that we knew to be wrong, because it would bias the results in Table III toward thylakoid proteins. The range of hydrophobicities of the proteins in Table III is illustrated in Fig. 5. The generalized order of elution from the reverse-phase HPLC column shown in Fig. 1 is presented below the GRAVY index (grand average of hydropathicity) to compare the relative elution time with the predicted hydrophobicity. The GRAVY statistic provides a reasonable indication of HPLC retention but tends to underestimate hydrophobicity if the protein has many buried charged residues, as in the case of the light-harvesting proteins, which have 15 charged amino acid residues within the bilayer (24). Retention of elements of secondary structure could also lead to higher retention times than predicted as polypeptide backbone polarity is minimized.

Our predicted mass calculations of thylakoid-associated proteins make use of the extensive body of literature incorporating their known primary structural information. Genes for thylakoid proteins are found in both nuclear and chloroplast genomes, and thus some members are trafficked across two biological membranes whereas others are synthesized directly into the thylakoid itself. N-terminal modifications are limited to removal of predictable transit sequences, sporadic removal of initiating formyl methionine residues, and phosphorylation/acetylation of some proteins. There is one example of C-terminal proteolytic trimming and, importantly, no glycosylation. Identities of over half of the detectable proteins were proposed by matching measured masses to minimally processed calculated masses, aided substantially by the rather limited range of post-translational modifications affecting thylakoid proteins. A single gene can provide multiple intact mass tags when post-translational modification is heterogeneous. An example is the D1 polypeptide, which has a major species corresponding to the translated protein processed at the N terminus by fMet removal, acetylation of Thr^2, and a C-terminal proteolytic event, as well as minor species where this product is phosphorylated or palmitoylated (see Fig. 2). Thus chloroplast *psbA* has at least three functionally significant gene products before further heterogeneity arising from oxidative modification is considered. The light-dependent turnover of D1 is well documented (25, 26); palmitoylation apparently plays a role in folding, assembly, or targeting of newly synthesized D1 to functional PS II in granal stacks (12), and N-terminal phosphorylation appears to modulate the turnover of the protein with respect to photodamage (27). Identifi-

TABLE III

TagIdent search of the A. thaliana genomes

TagIdent (ca.expasy.org/tools/tagident.html) was used to search Swiss-Prot and TrEMBL for translated genes from the three *A. thaliana* genomes. Each search was performed in the indicated mass ranges. The Swiss-Prot and TrEMBL entry names are used in the table. The searches resulted in 712 responses of which 133 are localized to the chloroplast (underlined in the table). 51 of the chloroplast-localized responses are for known proteins associated with the thylkoid membranes.

Mass range	TagIdent results from Swiss-Prot and TrEMBL (PI = 7, PI range = 10.00, OS = ARATH)
3000–5000	2SS1-ssu, 2SS2-ssu, 2SS3-ssu, 2SS4-ssu, MT1A, MT1B, MT1C, PETG, PETL, PSAI, PSBF, PSBI, PSBJ, PSBK, PSBL, PSBM, PSBN, PSBX, PSBY-1, PSBY-2, RL41, THG4
5000–7000	AFP1, AFP2, AFP3, IF1C, ITI2, ITI3, ITIF, KIN1, KIN2, PSAJ, RK32, RL39, RL40, THG1, THG2, THG3, YCF9
7000–9000	2SS4-lsu, ACP2, ACP3, ATP5, ATP9, ATPH, CLV3, DCAM, EC1, EC2, GAS1, GAS2, GAS3, IM08, ITI1, ITI6, MT2A, MT2B, PSAC, PSBH, RK28, RK33, RL38, RPBX, RS28, RS30, RUXG, S1FA, S61B
9000–11000	15E1, 2SS1-lsu, 2SS2-lsu, 2SS3-lsu, A9, ACBP, ACP1, ACPM, CH10, EM6, FER2, FER, GAS4, IM09, IM10, IM13, LTP2, LTP5, NLT1, NULM, PLAS, PLAT, POC1, POC2, PSAN, PSBE, PSBR, R27A, RK23, RL37, RLA1, RR15, RR16, RR17, RR19, RUXF, Y331, YC15
11000–13000	CWPX, CYC, DAD1, DAD2, ERH, NULC, RA2A, RA2B, RL44, RR14, RR18, SMT3, SR09, SUI1, T2AG, TBCA, THH1, U123, U136, U139, VAG1, VAG2, VAG3, Y103, Y150, Y20L
13000–15000	ATPE, DHX1, GCSH, GCSI, NU3C, NU3M, OLE3, PFD4, PR1, PRO1, PRO2, PRO3, PRO4, PRO5, R12A, R12B, R12C, R311, RBS1, RBS2, RBS3, RBS4, RK14, RK20, RL34, RPBY, RR13, RS1A, RS20, RS26, RT12, SODC, SR14, THF1, THH2, THH3, THH4, THM1, THM2, THM3, VATF, YCF3
15000–17000	CAL2, CAL4, CAL6, CY51, CY52, DSR5, EM1, FRDA, GRP7, GRP8, NDK1, PFD2, PFD5, PSQ1, PSQ2, RI21, R122, RBFA, RK16, RK24, RL23, RL27, RL2A, RR11, RR8, RS16, RS17, RS24, SODP, THM4, UBC9, UBCA, VATL
17000–19000	ABP1, ATP4, BCB, CDI8, COPT, CYP1, CYP2, DH18, DR1, F825, GC5L, HS11, HS12, HS13, HS21, MGN, N7BM, NDK3, NUGC, OLE1, PETD, R11A, R11B, RK21, RK22, RK9, RL12, RL21, RL24, RR7, RS15, RS18, RT07, SP18, TCH2, TCTP, TL16, TL17, UBC1, UBC2, UBC3, UBC7, UBCY
19000–21000	12S1, 12S2, AG31, AIG2, APT1, ARF1, ARF2, ARF3, AXI1, AXI2, AXI4, AXI5, AXIK, AXIS, CYP4, DH14, DHX2, GSHY, HEVL, H2SM, NDK2, NU6C, NUIC, NUKM, OM24, PSP1, R111, RL11, RL18, RM16, RPB7, RT19, UBC6, Y314
21000–23000	APT2, ARF4, ARF5, ATPF, ATPO, AXI3, AXI6, AXIJ, BAS1, CB24, CBP2, CD13, CHIC, CLPP, COP9, GRP, HL3B, MI25, NUGM, NUIM, OM21, OM23, PFD3, PR5, PSP2, R112, RAB7, RAC1, RAC2, RAC3, RAC4, RAC5, RERA, RERB, RHA1, RK15, RL9, RM05, RNS3, RT10, SARA, SARB, SODM, T215, TF21, TF22, TL30, UBC4, UBC5, UMPK, UPP, YCF4
23000–25000	14P, A494, AG12, ARA1, ARA2, ARA3, ATPI, BCCP, CAHC, CALL, CB21, CB22, CYB6, D101, ERF4, GTH1, GTH3, GTH4, GTH5, GTH6, GTH7, GTHB, GTHC, GTZ2, HL3A, HS2C, MAD2, MRAY, NU6M, OM22, OSL3, PIMT, PIST, PRCE, PSB1, RAN2, RB1A, RB1B, RB1C, REHY, RERC, RL10, RL13, RL15, RNS1, ROC1, RR4, SODF, SYBR, TLPH, V281, V282, YB87
25000–27000	AG11, AG14, AG17, AG19, AG1Y, AXI7, AXIB, AXIC, AXID, AXIG, AXIH, CF1, D102, EF1B, EF1C, ERD2, ERF2, GL1, GS27, GS28, GTXA, GTZ1, HIS2, HT14, IF6, IFE1, IM17, KAD, KAP1, KAP2, MPU1, NUKC, ORN PS51, PS52, PSA2, PSO1, PSO2, PUR3, RAN1, RAN2, RD19, RK4, RR2, RR3, SSRA, TIPG, TIPR, TPIS, VATE, Y136, Y278, YC22
27000–29000	12S2, 143E, 143K, 143L, 143P, 3MG, AG13, AGL1, AGL2, AGL4, AGL5, AGL6, AGL8, AG, APX1, AXIA, BI1, CAH2, CEMA, FABG, GCH2, GL2C, GL2M, H11, H12, HAT7, IDI1, MER5, MSRA, PROC, PS61, PS62, PS71, PS72, PSA3, PSA4, RAIL, RL2, RL7, RNS2, ROC2, ROC3, RPIA, RS6, RU2A, TIPA, TPIC, TPIS, VSP1, VSP2, Y350
29000–31000	12S1, 143C, 143H, 143M, 143N, 143O, 143U, AG15, AGL3, AGL9, AP1, ARA5, ATP3, AXII, CHIA, CNX3, COQ3, COX3, DH10, DH47, GBF4, HAT5, HAT9, HIS7, HT22, IPYR, KC2D, PCN1, PCN2, PIR2, PIR4, PS11, PS12, RK2, RL7A, RS2, RS3A, TL29, VATD, WC1A, WC1B, WC1C, WC2A, WC2B, WC2C, Y230
31000–33000	AOX2, ATH7, AX1B, AX1C, CHIT, CHMU, COPE, CTNS, CX32, CYF, D3H1, DCAM, E132, HAT1, HAT2, HSF4, IDI2, KC2B, KC2C, KC2E, PIR3, PSY, PUR5, QORL, RS31, RSP4, STS1, STX, SUCA, TH14, TRPC
33000–35000	ADT1, ADT2, ASPG, ATH4, ATH5, AX1A, CC2A, CYSK, ERF5, FSTL, GBF1, GGH, GGPP, HAT3, HSF6, IFRH, ISPE, MGHG, MDHI, MDHM, OM40, P2A1, P2A2, PIR1, PPX1, PPX2, PRSC, RL0A, RL0B, RL0C, RL5, SFR1, STS3, TF2B, URIC
35000–37000	ACC1, ATH6, ATP1, ATP2, AXI8, AXI9, CAPA, CC2B, CPPM, CYSL, D100, D111, DAP1, DAP2, DEGP, FABH, FN3X, G3PA, G3PC, GBLP, HBPB, HEM3, HKL2, IF32, MDHC, MDHD, ODPB, ORC6, OTC, P2A3, P2A4, P2A5, PERC, PERE, PORA, PORB, PORC, PP11, PP12, PP13, PP14, PP15, PP16, PP17, PP18, RD21, RP3A, RP3B, S17P, SPD1, STXK, TCH3, TRB1
37000–39000	ALF, ARG1, ARG2, CAD2, CCSA, CGD1, CYSN, DCUP, DMC1, DPNP, FLS1, GAE2, GAE3, GBF2, GLN1, HEM6, KPR1, LIPB, MTD1, MTD2, MTDH, NRL1, NRL2, NRL3, NRL4, P1, P2, PDA6, PIP, PSBA, PYRB, PYRC, RPOA, SPD2, YB30
39000–40000	CAD1, F16P, FPP1, FPP2, FTSZ, G3PB, GAE1, HSF7, KC21, KC22, KC23, KPPR, NU1C, ODPA, PFTA, PSBD, RD22, RIR2, YJ83

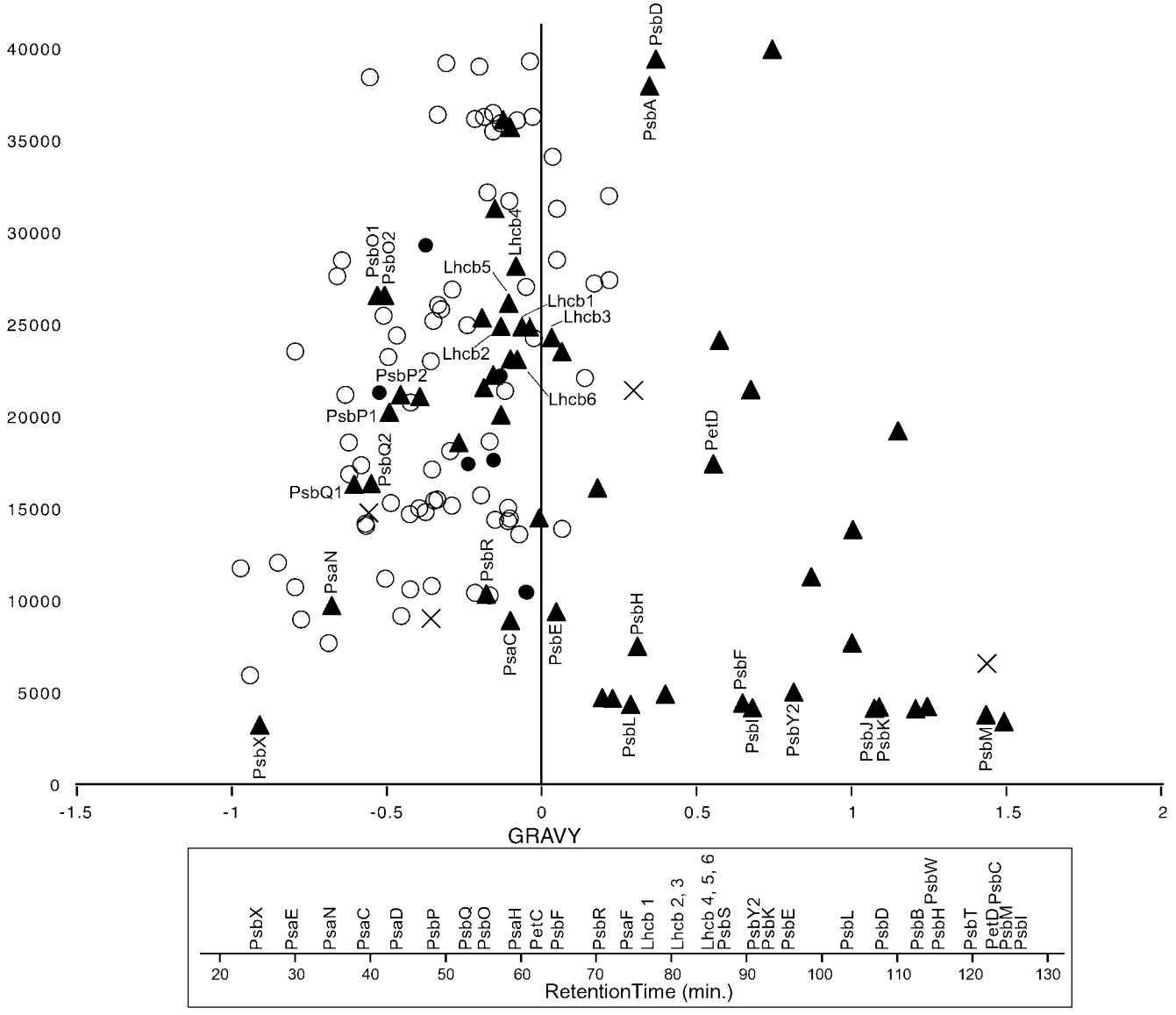

FIG. 5. **Predicted hydrophobicity of chloroplast proteins under 40,000 Da.** Proteins found by searching the ARATH subset of Swiss-Prot and TrEMBL with TagIdent (Table III) were analyzed using the GRAVY result from the ProtParam tool at ExPASy. The GRAVY index is shown (*abscissa*) *versus* the mass of each *A. thaliana* protein (*ordinate*). The *lower scale* shows the HPLC retention times of pea proteins from Fig. 1 with the approximate positions of each identified protein. The two scales were normalized to place PsbX and PsbM in the same relative positions. There are several membrane proteins that seem to elute later than the GRAVY index would indicate. It is possible that membrane proteins retain some secondary structure during reverse-phase HPLC and present a more hydrophobic surface than predicted for the denatured protein. *Filled triangles*, thylakoid-associated proteins (intrinsic and extrinsic); *open circles*, stromal proteins; *filled circles*, lumenal proteins; *X*, protein of unknown function encoded on the chloroplast genome (*ycf* genes).

cation of the IMTs assigned to D1 is practically unequivocal, because there are no other proteins of this size eluting with similar LC retention, and the assigned species closely match calculated masses for known modified forms of D1.

The presence of two phosphorylation sites on PsbH makes it possible for PsbH of green plants to exist in four different covalent states rather than two states as in cyanobacterial PsbH. Instead of a binary switch, the multiple phosphorylation states of PsbH in green plants allow a complex response to changing light environments. Simple light shift experiments such as those shown (Fig. 4), although demonstrating the photodynamic nature of phosphorylation of both PsbH sites, are not sufficient to distinguish the two sites, and more complex experiments will be required to discriminate the role of each site. The observation that PsbH has two phosphorylation sites demands reconsideration of the interpretation of many previous studies. For example, comparison of measured rates of phosphorylation/dephosphorylation of PsbH and LHC IIb

(Lhcb1) in greening pea chloroplasts showed that the rate of phosphorylation of PsbH is faster than that of LHC IIb, whereas the rate of dephosphorylation is much faster for LHC IIb (28). Similar rates were observed for the more rapid dephosphorylation of LHC IIb compared with PsbH in mature pea thylakoids (29). Several proteins are phosphorylated in PS II, but only the dephosphorylation kinetics of LHC IIb matched the kinetics of the decrease in low temperature (77 K) fluorescence emission ratios of PS I and LHC II (29). This has lead to a model where the phosphorylation of LHC IIb alone is correlated to the State 1-State 2 transition. However, different phosphorylation/dephosphorylation kinetics can be explained if there is a second site of phosphorylation on the PsbH protein (30). The demonstrated presence of a second phosphorylation site requires reconsideration of the role of PsbH in State 1-State 2 transitions and other light-induced physiological responses.

Not all IMTs are easy to assign, with some being more ambiguous and others remaining unassigned. In some cases modest alterations to a published gene sequence must be made to permit assignment. Such changes are based upon extreme conservation of specific residues across a wide range of sequences published for a variety of species.[3] Of course, all assignments are coincidental based upon match of intact mass to a calculated value, such that rigorous identification may be further required. This can be accomplished by splitting the effluent line between UV detector and mass spectrometer and running a proportion of the flow to a fraction collector for analysis by peptide mass fingerprints or sequence tags (LC-MS+; see Refs. 31 and 32). For example, LC-MS+ was used to identify a ninth component of chloroplast thylakoid cytochrome b_6f complexes as ferredoxin:NADP$^+$ oxidoreductase (33, 34). In another experiment, the spinach PsbF protein was sequenced by MS-MS[5] confirming that RNA editing results in a change of residue 26 from Ser to Phe (35), in agreement with MS studies of intact spinach and pea PsbF (3, 36). In the case of the IMT assigned to pea PsbH (7697.3 $\pm$ 1 Da), divergence from the calculated mass (7727.0 Da) is evident. Although peptide sequence obtained from MS-MS (not shown) shows that residue 18 is actually Val, rather than Ala as translated from the published DNA sequence (37), this change is insufficient to account for the mass difference between the observed and the calculated masses, and thus there are more changes to be found. Acetylation and phosphorylation of the N terminus of several thylakoid membrane proteins was identified in peptides isolated by immobilized metal ion chromatography using MS-MS (13, 38). Clearly complete annotation of genomic data will require detailed structural analysis by mass spectrometry.

One approach to monitor the thylakoid "phosphorylome" involves treatment of thylakoid membranes with trypsin, enrichment of phosphopeptides with immobilized metal ion chromatography, and their quantification by LC-MS-MS (17). However, yield of these peptides can be influenced by trypsin accessibility, and thylakoid protein conformation is known to be affected by phosphorylation (39, 40); furthermore, there is uncertainty about the efficiency of recovery of low abundance phosphopeptides by the immobilized metal ion chromatography procedure.[6] Therefore, care must be exercised in the interpretation of results obtained by this experimental method. An improvement would be to place such methodology downstream of the MS profile of the intact protein. For example, the oxidation of PsbH (see Fig. 4) would alter the mass such that a peptide may no longer be counted (unless specific incorporation of procedures to identify these species are used). Of course these detailed characterizations will be essential to clarify the relationship between membrane protein oxidation and phosphorylation.

The appearance of +32-Da adducts on phosphorylated forms of PsbH (Fig. 4*B*) would suggest oxidative addition of dioxygen via sulfone or endoperoxide formation, for example. Adducts of +32 Da were also observed when peptides derived from light-treated PS II reaction centers were analyzed (41) and contrast with +16-Da adducts that typically result from oxidation of methionine to its sulfoxide in aqueous buffers (6). Furthermore, it appears that the most highly phosphorylated forms of PsbH are also the most highly modified with the +32 adducts suggesting a link between the two phenomena. Detailed investigations of oxidative damage to proteins in biological membranes under physiological conditions are necessary.

CONCLUSION

Electrospray-ionization mass spectrometry, coupled with liquid chromatography, is being used (3, 4, 32–34, 41–44) to analyze intact thylakoid membrane proteins. LC-MS experiments provide an attractive means to monitor physiological changes in covalent status across the entire complement of thylakoid proteins and in subfractions from different membrane domains as a function of light and other stresses, providing significant benefits to functional genomics. Although it has been possible to propose identities of many thylakoid membrane proteins based upon coincidence of measured with calculated masses, it may be impractical to do so where there are much greater varieties of gene products (humans, for example) with more extensive modifications, especially glycosylation. However, the incorporation of fraction collection concomitant with intact mass measurements (LC-MS+) provides a convenient solution to the problem.

Acknowledgments—We gratefully acknowledge the advice from numerous collaborators and colleagues including Cameron Gunderson, Ernest Wright, H. Ronald Kaback, Johannes le Coutre, Eric Turk, John E. Walker, Ian Fearnley, James Barber, John F. Allen, Wolfgang Haenel, Werner Külbrandt, Fred W. McLafferty, Richard L. Stevens, Arvan Fluharty, and Guido Zamphigi.

[5] J. P. Whitelegge and L. Martin, III, unpublished data.

[6] R. H. Lee, personal communication.

* This work was supported in part by National Institutes of Health Research Grants DA 05010 (to C. J. Evans, principal investigator, and K. F. F.), NS 31271 (to A. L. Fluharty, principal investigator, K. F. F., and J. P. W.), AI 12601 (to M. Lovett, principal investigator, and J. P. W.), and NS 07171 (to A. D. Grinnell, principal investigator, and J. P. W.); by United States Department of Agriculture Grant 98-35100-6106 (to J. N. Nishio, principal investigator, S. M. G., and J. N. N.); by The Ford Foundation (to S. M. G.); and by equipment grants from the W. M. Keck Foundation and the Pasarow family. Some data are presented with assistance from the Kit Steinback Memorial Fund. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This paper is dedicated to J. Philip Thornber.

¶ To whom correspondence should be addressed: The Pasarow Mass Spectrometry Laboratory, Dept. of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095. Tel.: 310-206-7886; Fax: 310-206-2161; E-mail: jpw@chem.ucla.edu.

REFERENCES

1. Aebersold, R., Rist, B., and Gygi, S. P. (2000) Quantitative proteome analysis: methods and applications. *Ann. N. Y. Acad. Sci.* **919,** 33–47
2. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 9390–9395
3. Whitelegge, J. P., Gundersen, C., and Faull, K. F. (1998) Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci.* **7,** 1423–1430
4. Whitelegge, J. P., Faull, K. F., Gundersen, C., and Gómez, S. M. (1999) in *Photosynthesis: Mechanisms and Effects* (Garab, G., ed) Vol. V, pp. 4381–4384, Kluwer Academic Publishers, Dordrecht, The Netherlands
5. Whitelegge, J. P., le Coutre, J., Lee, J. C., Engel, C. K., Privé, G. G., Faull, K. F., and Kaback H. R. (1999) Toward the bilayer proteome, electrospray-ionization mass spectrometry of large intact transmembrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10695–10698
6. Whitelegge, J. P., Penn, B., To, T., Waring, A., Sherman, M., Stevens, R. L., Fluharty, C. B., Faull, K. F., and Fluharty, A. L. (2000) The effect of methionine oxidation upon the structure and function of the cerebroside-sulfate activator protein. *Protein Sci.* **9,** 1618–1630
7. le Coutre, J., Whitelegge, J. P., Gross, A., Turk, E., Wright, E. M., Kaback, H. R., and Faull K. F. (2000) Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry* **39,** 4237–4242
8. Whitelegge, J. P., and le Coutre, J. (2001) Proteomics: making sense of genomic information for drug discovery. *Am. J. Pharmacogenomics* **1,** 29–35
9. Turk, E., Kim, O., le Coutre, J., Whitelegge, J. P., Eskandari, S., Lam, J. T., Kreman, M., Zampighi, G., Faull, K. F., and Wright, E. M. (2000) Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters. *J. Biol. Chem.* **275,** 25711–25716
10. Peter, G. F., and Thornber, J. P. (1991) Biochemical composition and organization of higher plant photosystem II light-harvesting pigment proteins. *J. Biol. Chem.* **266,** 16745–16785
11. Whitelegge, J. P., Jewess, P., Pickering, M. G., Gerrish, C., Camilleri, P., and Bowyer, J. R. (1992) Sequence analysis of photoaffinity-labeled peptides derived by proteolysis of photosystem 2 reaction centers from thylakoid membranes treated with [^{14}C]azidoatrazine. *Eur. J. Biochem.* **207,** 1077–1084
12. Mattoo, A. K., and Edelman, M. (1987) Intramembrane translocation and post-translational palmitoylation of the chloroplast 32-kDa herbicide-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1497–1504
13. Michel, H., Hunt, D. F., Shabanowitz, J., and Bennett, J. (1988) Tandem mass spectrometry reveals that three photosystem II proteins of spinach chloroplasts contain *N*-acetyl-*O*-phosphothreonine at their NH_2 termini. *J. Biol. Chem.* **263,** 1123–1130
14. Takahashi, M., Shiraishi, T., and Asada, K. (1988) COOH-terminal residues of D1 and the 44-kDa CPa-2 at spinach photosystem II core complex. *FEBS Lett.* **240,** 6–8
15. Sharma, J., Panico, M., Barber, J., and Morris, H. R. (1997) Purification and determination of intact molecular mass by electrospray ionization mass spectrometry of the photosystem II reaction center subunits. *J. Biol. Chem.* **272,** 33153–33157
16. Gómez, S. M., Park, J. J., Zhu, J., Whitelegge, J. P., and Thornber, J. P. (1999) in *Photosynthesis: Mechanisms and Effects* (Garab, G., ed) Vol. I, pp. 353–356, Kluwer Academic Publishers, Dordrecht, The Netherlands
17. Vener, A. V., Harms, A., Sussman, M. R., and Vierstra, R. D. (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *J. Biol. Chem.* **276,** 6959–6966
18. Whitelegge, J. P. (1996) in *The Handbook of Photosynthesis* (Pessarakli, M., ed) pp. 241–256, Marcel Dekker, Inc., New York
19. Kieselbach, T., Bystedt, M., Hynds, P., Robinson, C., and Schroder, W. P. (2000) A peroxidase homologue and novel plastocyanin located by proteomics to the Arabidopsis chloroplast thylakoid lumen. *FEBS Lett.* **480,** 271–276
20. Kieselbach, T., Hagman, A., Andersson, B., and Schroder, W. P. (1998) The thylakoid lumen of chloroplasts: isolation and characterization. *J. Biol. Chem.* **273,** 6710–6716
21. 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
22. Berthold, D. A., Babcock, G. T., and Yocum, C. F. (1981) A highly resolved, oxygen-evolving photosystem 2 preparation from spinach thylakoid membranes. *FEBS Lett.* **134,** 231–234
23. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815
24. Külbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994) Atomic model of plant light-harvesting complex. *Nature* **367,** 614–621
25. Mattoo, A. K., Pick, U., Hoffman-Falk, H., and Edelman, M. (1981) The rapidly metabolized 32,000-dalton polypeptide of the chloroplast is the "proteinaceous shield" regulating photosystem II electron transport and mediating diuron herbicide sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 1572–1576
26. Prasil, O., Adir, N., and Ohad, I. (1992) in *Topics in Photosynthesis 11 (Photosystem II: Structure, Function and Molecular Biology)* (Barber, J., ed) pp. 295–348, Elsevier Science Publishers B.V., Amsterdam
27. Koivuniemi, A., Aro, E. M., and Andersson, B. (1995) Degradation of the D1 and D2 proteins of photosystem II in higher plants is regulated by reversible phosphorylation. *Biochemistry* **34,** 16022–16029
28. Black, M. T., Foyer, C. H., and Horton, P. (1984) An investigation into the ATP requirement for phosphorylation of thylakoid proteins and for the ATP-induced decrease in the yield of chlorophyll fluorescence in chloroplasts at different stages of development. *Biochim. Biophys. Acta* **767,** 557–562
29. Steinback, K. E., Bose, S., and Kyle, D. J. (1982) Phosphorylation of the light-harvesting chlorophyll protein regulates excitation energy distribution between photosystem II and photosystem I. *Arch. Biochem. Biophys.* **216,** 356–361
30. Sumner, E. J., Schmid, V. H., Bruns, B. U., and Schmidt, G. W. (1997) Requirement for the H phosphoprotein in photosystem II of *Chlamydomonas reinhardtii*. *Plant Physiol.* **113,** 1359–1368
31. Whitelegge, J. P. (1998) in *The Handbook of Plant and Crop Stress* (Pessarakli, M., ed) pp. 555–568, Marcel Dekker, Inc., New York
32. Whitelegge, J. P., Gómez, S. M., Nishio, J., and Faull, K. F. (2001) in *Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics*, *Chicago*, *May 28-June 1*, *2001*, Abstract WPH-151, American Society for Mass Spectrometry, Santa Fe, NM
33. Zhang, H., Whitelegge, J. P., and Cramer, W. A. (2001) Ferredoxin:NADP$^+$ oxidoreductase is a subunit of the chloroplast cytochrome b_6f complex. *J. Biol. Chem.* **276,** 38159–38165
34. Whitelegge, J. P., Zhang, H., and Cramer, W. A. (2001) *Proceedings of the 12th International Congress on Photosynthesis*, *Brisbane*, *August 18–23*, *2001*, Abstract S11-022, CSIRO Publishing, Collingwood, Victoria, Australia
35. Bock, R., Hagemann, R., Kössel, H., and Kudla, J. (1993) Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids-a new regulatory mechanism? *Mol. Gen. Genet.* **240,** 238–244
36. Sharma, J., Panico, M., Barber, J., and Morris, H. R. (1997) Characterization

of the low molecular weight photosystem II reaction center subunits and their light-induced modifications by mass spectrometry. *J. Biol. Chem.* **272,** 3935–3943
37. Lehmbeck, J., Stummann, B. M., and Henningsen, K. W. (1989) Nucleotide sequence of the 5.6-kbp *psbB* operon of pea chloroplast DNA. *Physiol. Plant.* **76,** 57–64
38. Michel, H., Griffin, P. R., Shabanowitz, J., Hunt, D. F., and Bennett, J. (1991) Tandem mass spectrometry identifies sites of three post-translational modifications of spinach light-harvesting chlorophyll protein II. Proteolytic cleavage, acetylation, and phosphorylation. *J. Biol. Chem.* **266,** 17584–17591
39. Nilsson, A., Stys, D., Drakenberg, T., Spangfort, M. D., Forsen, S., and Allen, J. F. (1997) Phosphorylation controls the three-dimensional structure of plant light-harvesting complex II. *J. Biol. Chem.* **272,** 18350–18357
40. Zer, H., Vink, M., Keren, N., Dilly-Hartwig, H. G., Paulsen, H., Herrmann, R. G., Andersson, B., and Ohad, I. I. (1999) Regulation of thylakoid protein phosphorylation at the substrate level: reversible light-induced conformational changes expose the phosphorylation site of the light-harvesting complex II. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8277–8282
41. Sharma, J., Panico, M., Shipton, C. A., Nilsson, F., Morris, H. R., and Barber, J. (1997) Primary structure characterization of the photosystem II D1 and D2 subunits. *J. Biol. Chem.* **272,** 33158–33166
42. Zheleva, D., Sharma, J., Panico, M., Morris, H. R., and Barber, J. (1998) Isolation and characterization of monomeric and dimeric CP47-reaction center photosystem II complexes. *J. Biol. Chem.* **273,** 16122–16127
43. Corradini, D., Huber, C. G., Timperio, A. M., and Zolla, L. (2000) Resolution and identification of the protein components of the photosystem II antenna system of higher plants by reverse-phase liquid chromatography with electrospray-mass spectrometric detection. *J. Chromatogr. A.* **886,** 111–121
44. Zolla, L., and Timperio, A. M. (2000) High performance liquid chromatography-electrospray mass spectrometry for the simultaneous resolution and identification of intrinsic thylakoid membrane proteins. *Proteins* **41,** 398–406