Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 67 (2006) 2431–2440
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Separation and identification of soybean leaf proteins by two-dimensional gel electrophoresis and mass spectrometry ☆

Chenping Xu [a], Wesley M. Garrett [b], Joseph Sullivan [a], Thomas J. Caperna [c], Savithiry Natarajan [d,*]

[a] Department of Natural Resource Sciences and Landscape Architecture, University of Maryland, College Park, MD 20742, USA
[b] US Department of Agriculture-Agricultural Research Service, Biotechnology and Germplasm Laboratory, Beltsville, MD 20705, USA
[c] US Department of Agriculture-Agricultural Research Service, Growth Biology Laboratory, Beltsville, MD 20705, USA
[d] US Department of Agriculture-Agricultural Research Service, Soybean Genomics and Improvement Laboratory, PSI, Beltsville, MD 20705, USA

Received 23 December 2005; received in revised form 17 February 2006

Abstract

To establish a proteomic reference map for soybean leaves, we separated and identified leaf proteins using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry (MS). Tryptic digests of 260 spots were subjected to peptide mass fingerprinting (PMF) by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS. Fifty-three of these protein spots were identified by searching NCBInr and SwissProt databases using the Mascot search engine. Sixty-seven spots that were not identified by MALDI-TOF-MS analysis were analyzed with liquid chromatography tandem mass spectrometry (LC-MS/MS), and 66 of these spots were identified by searching against the NCBInr, SwissProt and expressed sequence tag (EST) databases. We have identified a total of 71 unique proteins. The majority of the identified leaf proteins are involved in energy metabolism. The results indicate that 2D-PAGE, combined with MALDI-TOF-MS and LC-MS/MS, is a sensitive and powerful technique for separation and identification of soybean leaf proteins. A summary of the identified proteins and their putative functions is discussed.

Keywords: *Glycine max*; Soybean; Proteomics; Two-dimensional gel electrophoresis; MALDI-TOF MS; LC-MS/MS

1. Introduction

The two key steps in classical proteomics are the separation of proteins and their subsequent identification. In a standard approach, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry (MS) are combined. Two-dimensional PAGE, in which proteins are separated according to their isoelectric point (pI) in the first dimension and molecular weight (Mr) in the second dimension, is still the preferred separation technique of many researchers in the global and comparative analysis of proteins. MS has essentially replaced the classical technique of Edman degradation in protein identification because it is more sensitive, can deal with protein mixtures, and offers much higher throughput. There are two main approaches to MS protein identification. In peptide mass fingerprinting (PMF), the unknown protein is digested with a protease of known specificity such as trypsin. By determining the masses of the resulting peptides, a mass map

☆ Mention of trade name, proprietary product or vendor does not constitute a guarantee or warranty of the product by the US Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

* Corresponding author. Address: PSI-ARS-USDA, Soybean Genomics and Improvement Laboratory, 10300, Baltimore Avenue, Beltsville, MD 20705, USA. Tel.: +1 301 504 5258; fax: +1 301 504 5728.
E-mail address: natarajs@ba.ars.usda.gov (S. Natarajan).

0031-9422/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2006.09.002

or mass fingerprint can be obtained. This mass map is then compared with predicted mass maps of proteins within the database. The tandem mass spectrometric (MS/MS) method relies on fragmentation of individual peptides to obtain sequence information.

A number of plant proteomic studies have recently been published. Some focused on organelle or subcellular proteomes such as the chloroplast (Ferro et al., 2003; Lonosky et al., 2004), the mitochondria (Kruft et al., 2001; Bardel et al., 2002) or the ribosome (Yamaguchi et al., 2000, 2002), whereas others have focused on a specific tissue, such as *Arabidopsis* seeds (Gallardo et al., 2001), maize root (Chang et al., 2000) and maize leaves (Porubleva et al., 2001), soybean seed (Mooney and Thelen, 2004; Natarajan et al., 2005), pea leaves (Schiltz et al., 2004) and legume barrel medic roots (Mathesius et al., 2001). Large-scale projects to identify proteins from multiple tissues of the barrel medic (Watson et al., 2003), rice (Komatsu et al., 2004) and *Arabidopsis* (Giavalisco et al., 2005) have also been reported.

Soybean, *Glycine max* (*G. max*), provides an inexpensive source of protein for human food and for the animal industry and has been the dominant oilseed produced since the 1960s. So far, no protein reference map has been reported for soybean leaves. As a first step to study stress physiology of soybean, we separated and identified soybean leaf proteins from normal plants. We describe here the extraction and separation of soybean leaf proteins on 2D-PAGE gels and identification of proteins using both MALDI-TOF-MS and LC-MS/MS. The Mascot search engine was used to search against NCBInr, SwissProt and EST databases for protein identification.

2. Results and discussion

2.1. Separation of soybean leaf proteins

A previously described protein extraction protocol using acetone/trichloroacetic acid (TCA) precipitation was employed for the extraction of protein from soybean leaf (Natarajan et al., 2005). The proteins separated by 2D-PAGE were visualized by Coomassie brilliant blue (CBB) G-250 staining, which is MS compatible and allows reproducible protein detection. The resulting 2D-PAGE images, which were reproduced from four independent biological experiments, constitute the reference gel images. The simplicity of the protocol favors the reproducibility of the protein separation. A representative 2D-PAGE protein pattern of soybean leaf is presented in Fig. 1. The dynamic range of protein accumulation is very large; this is a problem for leaf proteomic analysis because the preponderance of ribulose bisphosphate carboxylase/oxygenase (Rubisco) masks the detection of other proteins (Wilson et al., 2002; Watson et al., 2003). However, in our system a significant number of proteins were clearly separated and identified despite the predominance of Rubisco in the central portion of the gels.

2.2. Identification of separated proteins

To qualitatively survey the proteins visualized by 2D-PAGE, a total of 260 protein spots were excised from the 2D-PAGE gels and digested with the trypsin. The peptide fragments were extracted and analyzed by MALDI-TOF-MS. Typically, high-quality MALDI-TOF-MS peptide

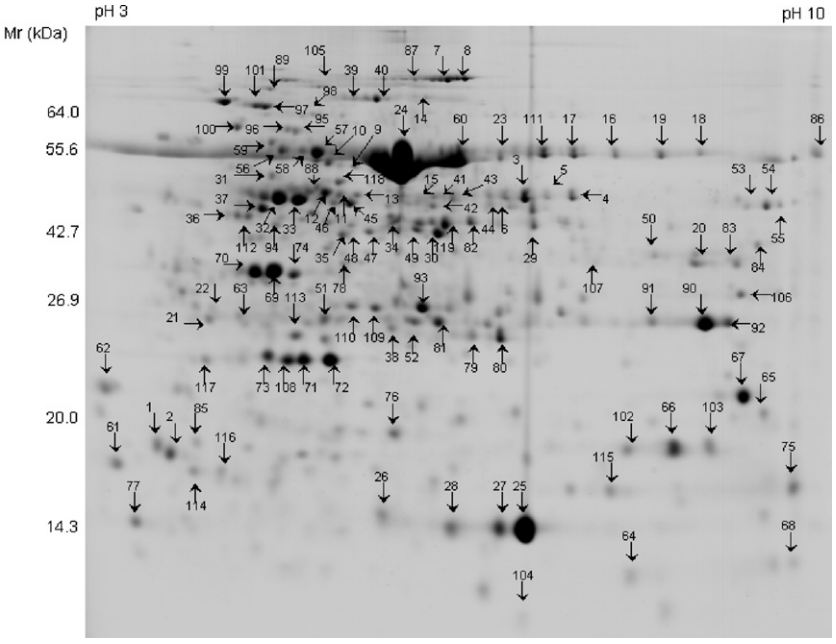

Fig. 1. Coomassie stained 2D-PAGE gel of separated soybean leaf proteins. Proteins were separated in the first dimension on a IPG strip pH 3.0–10.0 and in the second dimension on a 12.5% acrylamide SDS-gel. The numbered spots were identified and the derived data are presented in Table 1.

Table 1
Proteins identified from the soybean leaf by MALDI-TOF-MS and LC-MS/MS

SID	Protein identification [species]	T. Mr/pI	MO	PM	SC (%)	Acce. no.	Databases	ID Method
Functional category 1: Metabolism								
1	H-protein of glycine cleavage system [*Glycine soja*]	17744/4.59	65	2	10	gi|26045029	EST_others	LC-MS/MS
2	H-protein of glycine cleavage system [*Glycine soja*]	17744/4.59	145	3	22	gi|26045029	EST_others	LC-MS/MS
3	T-protein of the glycine cleavage system [*Pisum sativum*]	44661/8.79	131	3	7	gi|407475	NCBInr	LC-MS/MS
4	T-protein of the glycine cleavage system [*Pisum sativum*]	44661/8.79	506	10	25	gi|407475	NCBInr	LC-MS/MS
5	T-protein of glycine cleavage system, mitochondrial precursor [*Pisum sativum*]	44656/8.79	65	8	25	P49364	SwissProt	MALDI-TOF
6	T-protein of glycine cleavage system, mitochondrial precursor [*Pisum sativum*]	44661/8.79	151	3	7	gi|407475	NCBInr	LC-MS/MS
7	P-protein of glycine cleavage system, [*Arabidopsis thaliana*]	114672/6.18	463	9	10	gi|3413705	NCBInr	LC-MS/MS
8	P-protein of glycine cleavage system, [*Arabidopsis thaliana*]	114672/6.18	514	12	12	gi|3413705	NCBInr	LC-MS/MS
9	Alanine aminotransferase [*Arabidopsis thaliana*]	53780/6.49	164	4	9	gi|23297208	NCBInr	LC-MS/MS
10	Alanine aminotransferase [*Arabidopsis thaliana*]	53780/6.49	128	3	7	gi|23297208	NCBInr	LC-MS/MS
11	Glutamine synthetase precursor [*Glycine max*]	47948/6.73	83	12	35	gi|13877511	NCBInr	MALDI-TOF
12	Glutamine synthetase precursor [*Glycine max*]	47948/6.73	70	9	21	gi|13877511	NCBInr	MALDI-TOF
13	Glutamine synthetase precursor [*Glycine max*]	47948/6.73	407	12	14	gi|13877511	NCBInr	LC-MS/MS
14	Methionine synthase [*Glycine max*]	84401/5.93	143	24	36	gi|33325957	NCBInr	MALDI-TOF
15	Aspartate transaminase AAT5 precursor [*Glycine max*]	50725/7.16	81	14	36	gi|485495	NCBInr	MALDI-TOF
16	Glycine hydroxymethyltransferase [*Flaveria pringlei*]	57127/8.72	71	17	30	gi|437995	NCBInr	MALDI-TOF
17	Glycine hydroxymethyltransferase [*Flaveria pringlei*]	57068/8.8	234	4	10	gi|437997	NCBInr	LC-MS/MS
18	Glycine hydroxymethyltransferase, mitochondrial precursor[*Solanum tuberosum*]	57224/8.40	55	19	29	P50433	SwissProt	MALDI-TOF
19	Glycine hydroxymethyltransferase, mitochondrial precursor [*Arabidopsis thaliana*]	57534/8.4	157	2	5	Q9SZJ5	SwissProt	LC-MS/MS
20	Probable γ-glutamyl hydrolase [*Glycine max*]	37824/6.08	92	9	29	gi|7488702	NCBInr	MALDI-TOF
21	Ribose 5-phosphate isomerase [*Glycine max*]	19726/4.69	386	6	40	gi|15285625	EST_others	LC-MS/MS
22	Ribose 5-phosphate isomerase [*Glycine max*]	19726/4.69	118	2	15	gi|15285625	EST_others	LC-MS/MS
Functional category 2: Energy/pentose phosphate								
23	Rubisco large subunit [*Ophioglossum engelmannii*]	53034/5.96	104	15	33	gi|309636	NCBInr	MALDI-TOF
24	Rubisco large subunit [*Glycine max*]	52802/6.09	163	16	33	gi|3114769	NCBInr	MALDI-TOF
6	Ribulose-1,5-bisphosphate carboxylase [*Lechenaultia heteromera*]	43938/6.36	152	4	10	gi|1304320	NCBInr	LC-MS/MS
25	Rubisco small subunit rbcS2 [*Glycine max*]	20220/8.87	150	14	57	gi|10946377	NCBInr	MALDI-TOF
26	Rubisco small subunit rbcS2 [*Glycine max*]	20220/8.87	91	6	28	gi|10946377	NCBInr	MALDI-TOF
27	Rubisco small subunit rbcS2 [*Glycine max*]	20220/8.87	175	15	57	gi|10946377	NCBInr	MALDI-TOF
28	Rubisco small chain 4, chloroplast precursor [*Glycine max*]	20232/8.87	301	6	26	gi|132113	NCBInr	LC-MS/MS
29	Rubisco small chain 4, chloroplast precursor [*Glycine max*]	20232/8.87	94	10	56	gi|132113	NCBInr	MALDI-TOF
30	Rubisco activase precursor [*Datisca glomerata*]	41045/7.59	74	2	6	gi|3687652	NCBInr	LC-MS/MS
31	Rubisco activase [*Chenopodium quinoa*]	47925/6.56	386	6	18	gi|21950712	NCBInr	LC-MS/MS
32	Rubisco activase, chloroplast precursor [*Vigna radiata*]	48042/7.57	123	13	29	gi|10720249	NCBInr	MALDI-TOF
33	Rubisco activase, chloroplast precursor [*Vigna radiata*]	48042/7.57	73	9	22	gi|10720249	NCBInr	MALDI-TOF
34	Rubisco activase, chloroplast precursor[*Vigna radiata*]	48042/7.57	68	11	32	gi|10720249	NCBInr	MALDI-TOF
35	Rubisco activase β form precursor [*Deschampsia antarctica*]	47371/7.57	86	2	9	gi|32481063	NCBInr	LC-MS/MS
36	Sedoheptulose-1,7-bisphosphatase [*Spinacia oleracea*]	42568/5.87	88	2	5	gi|2529376	NCBInr	LC-MS/MS
112	Gm_ck31838 Soybean induced by Salicylic Acid (Sedoheptulose-1,7-bisphosphatase) [*Glycine max*]	23722/7.89	271	7	23	gi|31464866	EST_others	LC-MS/MS
37	Phosphoribulokinase [*Pisum sativum*]	39230/5.41	106	14	40	gi|1885326	NCBInr	MALDI-TOF
38	Ribulose-phosphate 3-epimerase [*Spinacia oleracea*]	30632/8.23	233	2	12	gi|3264788	NCBInr	LC-MS/MS
39	Transketolase [*Spinacia oleracea*]	80744/6.2	238	3	6	gi|2529342	NCBInr	LC-MS/MS
40	Transketolase [*Spinacia oleracea*]	80744/6.2	259	3	6	gi|2529342	NCBInr	LC-MS/MS
Functional category 2: Energy/glycolysis/glyoxylate cycle/gluconeogensis								
41	Glyceraldehyde-3-phosphate dehydrogenase [*Capsicum annuum*]	34126/6.34	204	3	11	gi|18072799	NCBInr	LC-MS/MS
42	Glyceraldehyde-3-phosphate dehydrogenase [*Capsicum annuum*]	34126/6.34	311	6	17	gi|18072799	NCBInr	LC-MS/MS
43	Glyceraldehyde-3-phosphate dehydrogenase [*Pisum sativum*]	43597/8.8	294	4	14	gi|12159	NCBInr	LC-MS/MS
44	Glyceraldehyde-3-phosphate dehydrogenase A-subunit precursor [*Nicotiana tabacum*]	42122/6.6	364	6	17	gi|170237	NCBInr	LC-MS/MS
45	Phosphoglycerate kinase [*Nicotiana tabacum*]	50317/8.48	116	10	28	gi|1161600	NCBInr	MALDI-TOF

(*continued on next page*)

Table 1 (*continued*)

SID	Protein identification [species]	T. Mr/pI	MO	PM	SC (%)	Acce. no.	Databases	ID Method
46	Phosphoglycerate kinase [*Nicotiana tabacum*]	50317/8.48	104	12	36	gi|1161600	NCBInr	MALDI-TOF
13	Phosphoglycerate kinase precursor [*Solanum tuberosum*]	50594/7.68	232	4	9	gi|3328122	NCBInr	LC-MS/MS
47	Homologous to plastidic aldolases [*Solanum tuberosum*]	38632/5.89	289	5	14	gi|1781348	NCBInr	LC-MS/MS
35	Plastidic aldolases [*Solanum tuberosum*]	38632/5.89	307	5	18	gi|1781348	NCBInr	LC-MS/MS
48	Fructose-bisphosphate aldolase, chloroplast precursor [*Spinacia oleracea*]	42727/6.85	153	2	6	P16096	SwissProt	LC-MS/MS
49	Malate dehydrogenase, cytoplasmic [*Mesembryanthemum crystallinum*]	35817/6.00	66	7	31	O24047	SwissProt	MALDI-TOF
50	Malate dehydrogenase 2, glyoxysomal [*Brassica napus*]	38043/8.14	214	2	9	gi|4995091	NCBInr	LC-MS/MS
51	Triosephosphate isomerase [*Fragaria* × *ananassa*]	33505/7.64	76	11	37	gi|7650502	NCBInr	MALDI-TOF
52	Triosephosphate isomerase [*Glycine max*]	27441/5.87	108	10	51	gi|48773765	NCBInr	MALDI-TOF
53	Glycolate oxidase, peroxisomal [*Spinacia oleracea*]	40317/9.16	261	6	12	gi|121530	NCBInr	LC-MS/MS
54	Glycolate oxidase [*Lens culinaris*]	40907/9.38	277	7	15	gi|228403	NCBInr	LC-MS/MS
55	Glycolate oxidase [*Mesembryanthemum crystallinum*]	40644/9.02	71	9	28	gi|1773330	NCBInr	MALDI-TOF
Functional category 2: Energy/electron transport								
56	ATPase β subunit [*Oryza sativa (japonica cultivar-group)*]	45265/5.26	94	12	35	gi|56784992	NCBInr	MALDI-TOF
57	ATPase β subunit [*Crossosoma californicum*]	52019/5.20	194	22	57	gi|14718020	NCBInr	MALDI-TOF
58	ATPase β subunit [*Platytheca verticellata*]	51184/5.07	147	21	50	gi|7708546	NCBInr	MALDI-TOF
59	ATPase α subunit [*Lotus corniculatus* var. *japonicus*]	55803/5.22	136	13	33	gi|13358984	NCBInr	MALDI-TOF
60	ATPase α subunit [*Phaseolus vulgaris*]	55595/6.51	90	14	31	gi|169318	NCBInr	MALDI-TOF
61	Ferredoxin I precursor [*Glycine max*]	18448/4.84	122	3	16	gi|5666556	EST_others	LC-MS/MS
62	Ferredoxin-like protein [*Glycine max*]	22101/4.45	131	2	13	gi|23731638	EST_others	LC-MS/MS
Functional category 2: Energy/photosynthesis								
63	Chlorophyll *a*/*b* binding protein (LHCP AB 180) [*Arabidopsis thaliana*]	25036/5.12	103	2	6	gi|16374	NCBInr	LC-MS/MS
64	PSI PsaN subunit precursor [*Zea mays*]	12841/8.37	75	2	20	gi|2981214	NCBInr	LC-MS/MS
65	PSI D2 subunit [*Nicotiana sylvestris*]	22467/9.78	106	2	13	gi|19748	NCBInr	LC-MS/MS
66	PSI reaction centre subunit IV A [*Glycine max*]	16368/9.08	143	4	13	gi|5606709	EST_others	LC-MS/MS
67	PSI reaction centre subunit D precursor [*Solanum tuberosum*]	22849/9.63	72	11	44	gi|34787117	NCBInr	MALDI-TOF
68	PSII 10 KD Polypeptide precursor [*Glycine max*]	13753/9.73	113	2	18	gi|607356	EST_others	LC-MS/MS
69	PSII Oxygen-evolving enhancer protein 1, chloroplast precursor [*Pisum sativum*]	35100/6.25	91	9	27	gi|20621	NCBInr	MALDI-TOF
70	PSII Oxygen-evolving enhancer protein 1, chloroplast precursor [*Pisum sativum*]	35100/6.25	66	7	23	P14226	SwissProt	MALDI-TOF
71	PSII Oxygen-evolving enhancer protein 2, chloroplast precursor [*Pisum sativum*]	28201/8.29	169	6	14	P16059	SwissProt	LC-MS/MS
72	PSII Oxygen-evolving enhancer protein 2, [*Solanum tuberosum*]	28158/8.27	120	3	8	gi|1771778	NCBInr	LC-MS/MS
73	PSII Oxygen-evolving enhancer protein 2 precursor [*Glycine max*]	19825/4.81	200	6	28	gi|16995778	EST_others	LC-MS/MS
74	PSII oxygen-evolving complex protein 3 [*Nicotiana tabacum*]	35377/5.89	85	9	31	gi|505482	NCBInr	MALDI-TOF
75	PSII oxygen-evolving complex protein 3 [*Lycopersicon esculentum*]	24557/9.64	85	3	15	gi|51457944	NCBInr	LC-MS/MS
76	Chloroplast Rieske FeS protein [*Pisum sativum*]	24683/8.63	190	6	16	gi|20832	NCBInr	LC-MS/MS
77	Plastocyanin [*Cucurbita pepo*]	10544/4.34	63	3	42	gi|130265	NCBInr	LC-MS/MS
78	Ferredoxin-NADP+ reductase [*Arabidopsis thaliana*]	40643/8.32	381	6	18	gi|20465661	NCBInr	LC-MS/MS
48	Quinone oxidoreductase-like protein, chloroplast precursor [*Arabidopsis thaliana*]	41132/8.46	234	4	8	Q9ZUC1	SwissProt	LC-MS/MS
79	Carbonic anhydrase [*Vigna radiata*]	35804/7.59	305	8	20	gi|8954289	NCBInr	LC-MS/MS
80	Carbonic anhydrase [*Vigna radiata*]	35804/7.59	292	8	25	gi|8954289	NCBInr	LC-MS/MS
81	Carbonic anhydrase [*Vigna radiata*]	35804/7.59	267	7	14	gi|8954289	NCBInr	LC-MS/MS
Functional category 4: Transcription								
82	Chloroplast mRNA-binding protein CSP41 precursor [*Glycine max*]	22707/5.31	169	6	21	gi|7639890	EST_others	LC-MS/MS
119	Putative RNA-binding protein [*Arabidopsis thaliana*]	42303/7.71	241	4	12	gi|3850621	NCBInr	LC-MS/MS
Functional category 5: Protein synthesis								
83	30S ribosomal protein S5 [*Arabidopsis thaliana*]	32682/8.99	128	2	8	gi|21593322	NCBInr	LC-MS/MS
84	50S Ribosomal protein L1, chloroplast precursor [*Glycine soja*]	29084/9.43	119	2	9	gi|26046597	NCBInr	LC-MS/MS

Table 1 (*continued*)

SID	Protein identification [species]	T. Mr/pI	MO	PM	SC (%)	Acce. no.	Databases	ID Method
85	50S Ribosomal protein L12 [*Glycine max*]	16769/4.64	467	6	53	gi|22927916	EST_others	LC-MS/MS
86	Elongation factor-1 (EF-1a) [*Glycine max*]	49689/9.14	264	7	17	gi|18765	NCBInr	LC-MS/MS
87	Elongation factor 2 (EF-2) [*Beta vulgaris*]	94708/5.93	56	14	19	O23755	SwissProt	MALDI-TOF
88	Translation elongation factor-TU [*Glycine max*]	52177/6.21	170	22	55	gi|18776	NCBInr	MALDI-TOF
89	Translation elongation factor G [*Glycine max*]	77866/5.04	465	9	14	gi|402753	NCBInr	LC-MS/MS
Functional category 6: Protein destination and storage								
90	Stem 28 kDa protein [*Glycine max*]	29218/8.75	74	8	29	gi|169898	NCBInr	MALDI-TOF
91	Stem 28 kDa protein [*Glycine max*]	29218/8.75	82	9	33	gi|169898	NCBInr	MALDI-TOF
92	Stem 28 kDa protein precursor [*Glycine max*]	29218/8.75	61	7	30	P15490	SwissProt	MALDI-TOF
93	Vegetative storage protein, precursor [*Glycine max*]	29433/6.72	217	4	16	gi|72303	NCBInr	LC-MS/MS
94	PS II stability/assembly factor HCF136, chloroplast precursor [*Arabidopsis thaliana*]	44133/6.79	61	10	28	O82660	SwissProt	MALDI-TOF
95	Chaperonin precursor [*Pisum sativum*]	63287/5.85	105	10	24	gi|806808	NCBInr	MALDI-TOF
96	Chaperonin precursor [*Pisum sativum*]	63287/5.85	110	10	24	gi|806808	NCBInr	MALDI-TOF
113	Gm_ck32525 Soybean induced by SA (chaperonin 2)[*Glycine max*]	24022/5.44	249	4	33	gi|31465298	EST_others	LC-MS/MS
97	Endoplasmic reticulum HSC70-cognate binding protein precursor [*Glycine max*]	73822/5.15	88	15	23	gi|2642238	NCBInr	MALDI-TOF
98	70 kDa heat shock protein [*Phaseolus vulgaris*]	72721/5.95	87	19	27	gi|22636	NCBInr	MALDI-TOF
99	Heat shock protein 70 [*Cucumis sativus*]	75480/5.15	164	16	25	gi|1143427	NCBInr	MALDI-TOF
100	Chaperonin groEL [*Ricinus communis*]	52461/4.77	127	12	32	gi|72958	NCBInr	MALDI-TOF
101	Endoplasmic reticulum HSC70-cognate binding protein precursor [*Glycine max*]	73822/5.15	92	14	21	gi|2642238	NCBInr	MALDI-TOF
76	Cyclophilin, chloroplast precursor [*Oryza sativa (japonica cultivar-group)*]	25273/8.05	123	3	23	gi|34902534	NCBInr	LC-MS/MS
102	Cyclophilin [*Phaseolus vulgaris*]	18376/8.36	265	7	33	gi|829119	NCBInr	LC-MS/MS
103	Cyclophilin [*Glycine max*]	18395/8.7	174	2	15	gi|17981611	NCBInr	LC-MS/MS
104	Polyubiquitin 1 [*Phaseolus vulgaris*]	6547/6.51	67	8	63	gi|33327284	NCBInr	MALDI-TOF
105	Endopeptidase Clp ATP-binding chain cd4B [*Lycopersicon esculentum*]	102463/5.86	102	29	30	gi|9758239	NCBInr	MALDI-TOF
Functional category 7: Transporters								
106	Voltage-dependent anion-selective channel (VDAC1.2) [*Lotus corniculatus* var. *japonicus*]	29696/8.57	93	2	8	gi|36957183	NCBInr	LC-MS/MS
Functional category 10: Signal transduction								
63	Harpin binding protein 1 [*Glycine max*]	28471/7.88	131	3	10	gi|38679315	NCBInr	LC-MS/MS
107	Guanine nucleotide-binding protein β subunit-like protein [*Glycine max*]	35985/7.62	70	11	40	gi|1256608	NCBInr	MALDI-TOF
108	(P21_SOYBN) P21 protein [*Glycine max*]	22365/4.84	146	3	16	P25096	SwissProt	LC-MS/MS
Functional category 11: Disease/defence								
109	Ascorbate peroxidase 2 [*Glycine max*]	27180/5.65	154	12	58	gi|1336082	NCBInr	MALDI-TOF
110	Ascorbate peroxidase 2 [*Glycine max*]	27180/5.65	71	8	41	gi|1336082	NCBInr	MALDI-TOF
111	Catalase [*Glycine max*]	57043/6.80	123	15	32	gi|2661021	NCBInr	MALDI-TOF
114	Stress-induced protein SAM22 (Starvation-associated message 22) [*Glycine max*]	16762/4.69	282	4	31	P26987	SwissProt	LC-MS/MS
115	PR1A precursor [*Glycine max*]	18108/8.20	221	4	25	gi|13479525	EST_others	LC-MS/MS
116	Peroxiredoxin [*Hyacinthus orientalis*]	14120/5.43	98	2	21	gi|42565527	NCBInr	LC-MS/MS
117	Peroxiredoxin [*Phaseolus vulgaris*]	28776/5.17	232	6	25	gi|11558244	NCBInr	LC-MS/MS
Functional category 20: Secondary metabolism								
118	1-Deoxy-D-xylulose 5-phosphate reductoisomerase [*Pueraria montana* var. *lobata*]	50808/5.83	87	12	28	gi|35187000	NCBInr	MALDI-TOF
Functional category 12: Unclear classification								
116	Hypothetical protein [*Oryza sativa (japonica cultivar-group)*]	24917/7.49	95	2	9	gi|22165076	NCBInr	LC-MS/MS

The data were classified according to protein function described by Bevan et al. (1998) and included an assigned spot number (Fig. 1). SID: Spot ID; T. Mr/pI: theoretical molecular weight and pI; MO: MOWSE score; PM: the number of peptides matched; SC: the percentage of sequence coverage; Acce. no.: Accession number; ID method: identification method. The assigned protein of the best matched was given with the species in which it has been identified and its accession number.

mass maps were obtained. Of the 260 protein spots processed, 53 proteins were successfully identified by querying NCBInr and SwissProt databases using the Mascot search engine. The results are listed in Table 1. Data in Table 1 include an assigned protein spot number, theoretical pI and Mr, protein identity and its original species, number of peptides matched, percentage sequence coverage, MOWSE score, database searched, identification method and accession number of the best match. Although MALDI-TOF-MS analysis is easy to automate and allows high throughput analysis, protein identification relies solely on the accurate matching of the peptide mass, and it is very difficult to correctly identify proteins if there is a protein mixture. Also the database must contain enough of the protein sequence to compare with the experimentally derived mass map.

LC-MS/MS analysis of 67 spots that were not identified by MALDI-TOF MS allowed the identification of 66 spots, searching against the NCBInr, SwissProt and EST-others databases. The identifications are listed in Table 1. The efficiency of the identification is higher than seen in a proteomic study of rice where 77% of proteins were identified (Lin et al., 2005). An example of an LC-MS/MS spectrum obtained from spot 4 is presented in Fig. 2. The tandem mass spectrometric method is technically more complex and less scalable than MALDI fingerprinting. Its main advantage is that sequence information derived from several peptides is much more specific for the protein than a list of peptide masses. A short region of local identity, which spans two or more consecutive tryptic cleavage sites may enable identification, even when the remainder of the sequence is divergent (Pandey and Mann, 2000; Newton et al., 2004). In addition, the fragmentation data can be used to search nucleotide databases such as db EST as well as the protein databases. Most of the 119 identified protein spots were identified by searching against the NCBInr database and contained only one protein, seven spots contained two different proteins (spot 6, 13, 35, 48, 63, 76, and 116). Fourteen spots were identified from EST_others database, and 13 spots were identified from the SwissProt database. Only 35% of the identifications came from *G. max* or *Glycine soja* (*G. soja*) species.

The theoretical and experimental Mrs and pIs matched closely for 72 of the 119 identified spots. We used the BLAST tool in ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (http://us.expasy.org/) for annotation and found that many of the proteins identified have a signal peptide. We calculated the theoretical pI/Mr of these proteins without the signal sequence using the compute pI/Mr tool and found that additional 23 spots have matched Mrs and pIs. For example, the identification for spot 94 is photosystem (PS) II stability/assembly factor HCF136 precursor that has 403 amino acids, and pI/Mr of 6.79/44133. The first 78 amino acids is signal peptide, so the theoretical pI/Mr of the rest peptide is 5.15/35843, which closely

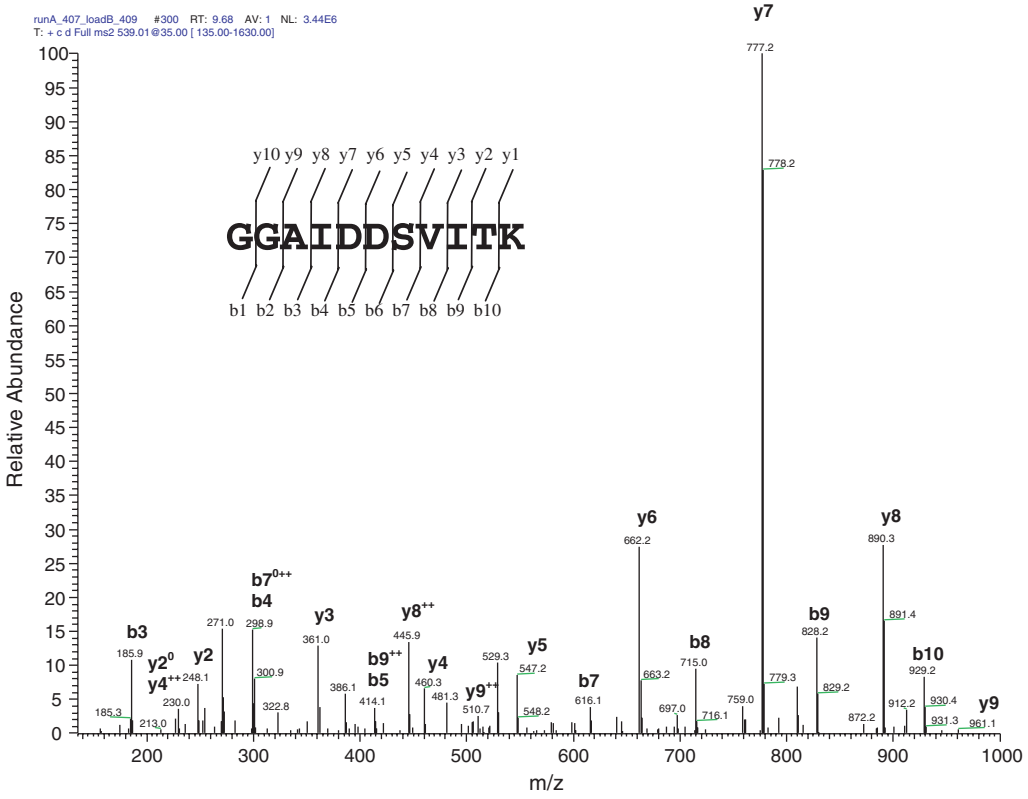

Fig. 2. MS/MS spectrum of one tryptic peptide of spot 4. The protein was identified as T-protein of the glycine cleavage system on the basis of finding 10 unique tryptic peptides, eight with significant ion score, and total protein coverage of 25%.

matches with our observed values. The pI/Mr discrepancy for the other spots might be due to different amino acid sequence from different species (spot 6, 79, 80, 81, 106, and 116), amino acid sequence derived from an EST that may not include complete or accurate sequence (spot 21, 22, 82 and 112), or co- and/or post-translational modification (spot 13, 29, and 86).

Multiple spots for a single protein are commonly found on 2D gels (Sarnighausen et al., 2004; Giavalisco et al., 2005). In this study 30 proteins have multiple spots: T-protein of glycine cleavage system (spot 3–6); Rubisco both large (spot 6, 23, and 24) and small (spot 25–29) subunits; Rubisco activase (spot 30–35); stem 28 kDa protein (spot 90–93); and many others. Several factors may be responsible for this phenomenon. The migration of proteins on a 2D-PAGE gel is very sensitive to small structural differences. These spots might be different isoforms derived from different genes of a multigene family. The complex genome of soybean is expected to contain multiple copies of many genes, and the distinct biophysical properties might be due to amino acid sequence differences in the different isoforms. Alternatively, one gene product may undergo different co- and/or post-translational modifications that affect its pI or/and Mr. The multiple spots corresponding to one protein could also be a consequence of artificial modification of proteins, such as carbamylation, during the extraction or separation procedure (Berven et al., 2003). However, appropriate precautions were used to prevent artificial modifications and the multiple spots are highly reproducible. Therefore, it is unlikely that the multiple spots in this study are artifacts of protein damage during sample preparation. Taking into account the multiplicity of the spots, we identified a total of 71 unique proteins on our gel.

Based on bioinformatic protein sequence analysis, proteins with at least one transmembrane-spanning domain constitute approximately 20% of all proteins in eukaryotic genomes (Wallin and von Heijne, 1998; Stevens and Arkin, 2000). These proteins are often underrepresented on 2D-PAGE gels due to the tendency of the hydrophobic transmembrane regions to cause the proteins to precipitate, mainly during isoelectric focusing (IEF) (Molloy et al., 1998; Santoni et al., 2000). The observation of plant proteins in 2D-PAGE relative to their general average hydropathicity score has been discussed (Millar et al., 2001). Most proteins identified in this study were expected to be soluble proteins, given that TCA precipitation results in the loss of integral membrane proteins (Wang et al., 2000). Nevertheless, we found some spots that were identified as integral or peripheral membrane proteins, such as chlorophyll *a*/*b* binding protein (spot 63), oxygen-evolving enhancer protein of PS II (spot 69–75), and subunits of the PS I complex (spot 64–67). These light-harvesting complexes of PS I and PS II are highly abundant in the thylakoid membranes of plant chloroplasts (Gomez et al., 2000). Therefore, a small fraction of these proteins was extracted during the sample preparation from the whole leaves.

2.3. Functional distribution of identified proteins

Identified proteins were classified according to their functions in the categories described by Bevan et al. (1998). In our study, more than 50% of the identified protein spots are involved in energy metabolism (Fig. 3). This category was divided into four subcategories: the pentose phosphate pathway, the glycolysis pathway/glyoxylate cycle/gluconeogenesis, electron transport, and photosynthesis. The most abundant proteins were also included in the energy category, and most of these proteins appeared as multiple spots. The main function of a plant leaf is energy harvesting, conversion, and storage. Therefore, it is not surprising that a significant number of abundant proteins in the leaf proteome are involved in energy metabolism. The identified proteins involved in photosynthetic electronic transport were: subunits of PS I (spot 64–67) and PS II (spot 63, 68–75); Rieske FeS protein (spot 76); plastocyanin (spot 77); and ferredoxin (spot 61 and 62). Several enzymes responsible for carbon metabolism were also identified: rubisco (spot 6, and 23–29); malate dehydrogenase (spot 49 and 50); sedoheptulose-1,7-biphosphatase (spot 36 and 112); phosphoglycerate kinase (spot 13, 45, and 46); glyceraldehyde-3-phosphate dehydrogenase (spot 41–44); triosephosphate isomerase (spot 51 and 52); and transketolase (spot 39 and 40). Rubisco is the primary enzyme in photosynthetic carbon fixation and the likely rate-limiting factor for photosynthesis under light-saturated conditions and atmospheric CO_2 pressure (Makino et al., 1985). The ATPase consists of two parts; a hydrophobic membrane-bound portion called CF_0, and a soluble portion that sticks out into the stroma called CF_1. CF_1 consists of 5 different subunits: α-, β-, γ-, δ-, and ε-units (Taiz and Zeiger, 2002). Only the soluble α- (spot 59 and

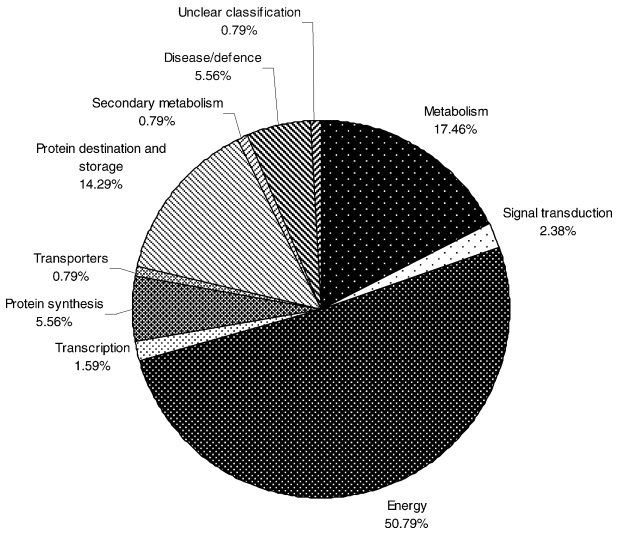

Fig. 3. Assignment of the identified proteins to functional categories using the classification described by Bevan et al. (1998). A total of 119 spots representing 71 different proteins were classified. If a spot contained two proteins it was counted twice.

60) and β- (spot 56–58) subunits of the ATPase complex were present on our 2D-PAGE gels.

About 17% of the identified protein spots are involved in general metabolism. Most of these are involved in amino acid metabolism: glutamine synthetase (spot 11–13); glycine (serine) hydroxymethyltransferase (spot 16–19); alanine aminotransferase (spot 9 and 10); methionine synthase (spot 14); aspartate transaminase (spot 15); P- (spot 7 and 8), and H- (spot 1 and 2) and T-protein (spot 3–6) of the glycine cleavage system. Glycine (serine) hydroxymethyltransferase (spot 16–19) catalyzes the interconversion of serine and glycine. It is a key enzyme in the biosynthesis of purines, lipids, hormones and other compounds (Kopriva and Bauwe, 1995). The glycine cleavage system catalyzes the degradation of glycine and is composed of four proteins: P, T, L and H proteins (Bourguignon et al., 1993). Other identified proteins belong in the protein destination and storage category: chaperonin (spot 95, 96, 100 and 113), HSP/HSC (heat shock protein/heat shock cognate) 70 and associated co-chaperones (spot 97, 98, 99, and 101), stem 28 kDa protein (spot 90–93), cyclophilin (spot 76, 102, and 103), endopeptidase Clp (spot 105) and polyubiquitin (spot 104). HSPs are associated with protein folding, protein translocation across membranes, assembly of oligomeric proteins, modulation of receptor activities, mRNA protection, prevention of enzyme denaturation and their stress-induced aggregation, and with post-stress ubiquitin and chaperonin-aided repair. Based on these functions, HSPs have been termed ''molecular chaperones'' (Georgopoulos and Welch, 1993; Leone et al., 2000). Stem 28 kDa glycoprotein is also known as vegetative storage protein A. It may function as a somatic storage protein during early seedling development and mainly accumulates in the stem of developing seedlings (Mason et al., 1988).

Additional identified protein spots are included in disease/defence category: ascorbate peroxidase (spot 109 and 110); catalase (spot 111); stress-induced protein SAM22 (spot 114); peroxiredoxin (spot 116 and 117); and PR1 (spot 115). Three signal transduction proteins were detected in our study: harpin-binding protein 1 (spot 63); P21 protein (spot 108); and G protein (spot 107). Other identified proteins are involved in protein synthesis (ribosomal protein, spot 83–85; elongation factor, spot 86–89), in transcription (RNA binding protein, spot 119 and 82), in ion transport (voltage-dependent anion-selective channel, spot 106), and in secondary metabolism (1-deoxy-D-xylulose 5-phosphate reductoisomerase, spot 118). Only one identified protein has unknown function (spot 116). None of the identified proteins are in the cell growth/division, intracellular traffic, cell structure or transposon categories.

3. Conclusion

We separated soybean leaf proteins using 2D-PAGE and identified 119 protein spots with MALDI-TOF MS and LC-MS/MS. The broad dynamic range of protein expression is one of the major difficulties in separation of soybean leaf proteins by 2D-PAGE. LC-MS/MS is a more powerful and sensitive way to obtain positive identifications from 2D-PAGE spots, although it is more laborious and more difficult to automate than MALDI-TOF-MS analysis. Future studies of leaf physiology will benefit from this proteome reference map of soybean leaf.

4. Experimental

4.1. Plant material

The seeds of soybean cultivar Clark were grown on horticultural vermiculite in 1-L pots in the greenhouse at the University of Maryland, College Park, MD. Plants were transferred to the field site at the USDA, Beltsville Southfarm just before the cotyledons emerged, and grown under full sunlight. Plants were watered and rotated daily, and fertilized with Hoagland solution every 3 days. Primary leaves were harvested when they were 12 days old. Four biological replicate samples were used for protein extraction and 2D-PAGE analysis. Samples were immediately frozen in liquid nitrogen, and stored at −80 °C prior to analysis.

4.2. Protein extraction from leaves

Frozen sample was ground in a mortar with liquid nitrogen and incubated with 10% TCA and 0.07% 2-mercaptoethanol in acetone for 1 h at −20 °C (Natarajan et al., 2005). The precipitated proteins were pelleted and washed with ice-cold acetone containing 0.07% 2-mercaptoethanol to remove pigments and lipids until the supernatant was colorless. The pellet was vacuum dried, resuspended in resolubilization solution (9 M urea, 1% CHAPS, 1% DTT, 1% pharmalyte) and sonicated to extract proteins. Insoluble tissue was removed by centrifugation at 21,000*g* for 30 min. Protein concentration was determined according to Bradford (1976) using a commercial dye reagent (Bio-Rad Laboratories, Hercules, CA) with bovine serum albumin as a standard.

4.3. 2D-PAGE

An IPGPhor apparatus (GE Healthcare, Piscataway, NJ) was used for IEF with immobilized pH gradient (IPG) strips (pH 3.0–10.0, linear gradient, 13 cm). The IPG strips were rehydrated 12 h with 250μL rehydration buffer (8 M urea, 2% CHAPS, 0.5% pharmalyte, 0.002% bromophenol blue) containing 350 μg proteins. The voltage settings for IEF was 500 V for 1 h, 1000 V for 1 h, 5000 V for 1 h, and 8000 V to a total 46.86 kVh. Following electrophoresis, the protein in the strips was denatured with equilibration buffer (50 mM Tris–HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, 1% DTT)

and then incubated with the same buffer containing 2.5% iodoacetamide instead of DTT for 30 min at room temperature. The second dimension electrophoresis was performed on a 12.5% gel using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare, Piscataway, NJ). The gels were stained with CBB G-250 (Newsholme et al., 2000) and scanned using a Personal Densitometer SI (GE Healthcare, Piscataway, NJ).

4.4. In-gel digestion of protein spots

Protein digestion was performed as described previously (Natarajan et al., 2005). Spots were excised from the stained gel and washed with CH_3CN:H_2O (1:1, v/v) containing 25 mM ammonium bicarbonate to remove the dye. The gel plug was dehydrated with 100% acetonitrile, and was dried under vacuum and incubated overnight at 37 °C with 20 μL of 10 μg/mL porcine trypsin in 20 mM ammonium bicarbonate. The resulting tryptic fragments were eluted by diffusion into CH_3CN:H_2O (1:1, v/v) and 0.5% trifluoroacetic acid. A sonic bath was used to facilitate the diffusion. The extract was vacuum dried and the pellet was dissolved in CH_3CN:H_2O (1:1, v/v) and 0.1% trifluoroacetic acid.

4.5. Mass spectrometry

For PMF a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA) operated in positive ion reflector mode was used to analyze tryptic peptides. Samples were co-crystallized with α-cyanohydroxycinnamic acid (CHCA) matrix, and spectra were acquired with 50 shots of a 337 nm Nitrogen Laser operating at 20 Hz. Spectra were calibrated using the trypsin autolysis peaks at m/z 842.51 and 2.211.10 as internal standards. For MS/MS a Thermo Finnigan LCQ Deca XP plus Ion Trap mass spectrometer was used to analyze proteins that were not positively identified by MALDI-TOF-MS. Peptides were separated on a reverse phase column using a 30 min gradient of 5–60% acetonitrile in water with 0.1% formic acid. The instrument was operated with a duty cycle that acquired MS/MS spectra on the three most abundant ions identified by a survey scan from 300 to 2000 Da. Dynamic exclusion was employed to prevent the continuous analysis of the same ions. Once two MS/MS spectra of any given ion had been acquired, the parent mass was placed on an exclusion list for the duration of 1.5 min. The raw data were processed by Sequest to generate DTA files for database searching. The merge.pl script from Matrix Science was used to convert multiple Sequest DTA files into a single mascot generic file suitable for searching in Mascot.

4.6. Data analysis

Protein identification was performed using the Mascot search engine (http://www.matrixscience.com), which uses a probability based scoring system (Perkins et al., 1999). NCBI non-redundant and SwissProt databases were selected as the primary databases to be searched. For LC-MS/MS, if the primary databases did not yield identity, the ''EST_others'' database was queried. The parameters for database searches with MALDI-TOF PMF data and with MS/MS spectra were set as before (Natarajan et al., 2005). For MALDI-TOF-MS data to qualify as a positive identification, a protein's score had to equal or exceed the minimum significant score of 64 for NCBInr or 55 for SwissProt database searching. Positive identifications of proteins by MS/MS analysis required a minimum of two unique peptides, with at least one peptide having a significant ion score.

Acknowledgements

The authors thank Dr. W. Kenworthy for providing soybean seeds and Dr. J. Slovin and Dr. N.C. Talbot for their critical reviews of this manuscript. Funding for this research was provided by ARS Project 1275-21000-15600D.

References

Bardel, J., Louwagie, M., Jaquinod, M., Jourdain, A., Luche, S., Rabilloud, T., Macherel, D., Garin, J., Bourguignon, J., 2002. A survey of the plant mitochondrial proteome in relation to development. Proteomics 2, 880–898.

Berven, F.S., Karlsen, O.A., Murrell, J.C., Jensen, H.B., 2003. Multiple polypeptide forms observed in two-dimensional gels of Methylococcus capsulatus (Bath) polypeptides are generated during the separation procedure. Electrophoresis 24, 757–761.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirske, W., Van Staveren, M., Stiekema, W., 1998. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391, 485–488.

Bourguignon, J., Vauclare, P., Merand, V., Forest, E., Neuburger, M., Douce, R., 1993. Glycine decarboxylase complex from higher plants. Molecular cloning, tissue distribution and mass spectrometry analyses of the T protein. Eur. J. Biochem. 217, 377–386.

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

Chang, W., Huang, L., Shen, M., Webster, C., Burlingame, A.L., Roberts, J., 2000. Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment and identification of proteins by mass spectrometry. Plant Physiol. 122, 295–317.

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

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vanderkerckhove, J., Job, D., 2001. Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiol. 126, 835–848.

Georgopoulos, C., Welch, W.J., 1993. Role of the major heat shock proteins as molecular chaperones. Annu. Rev. Cell Biol. 9, 601–634.

Giavalisco, P., Nordhoff, E., Kreitler, T., Kloppel, K., Lehrach, H., Klose, J., Gobom, J., 2005. Proteome analysis of *Arabidopsis thaliana* by

two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. Proteomics 5, 1902–1913.

Gomez, S., Chitins, P.R., Yunus, M., Pathre, U., Mohanty, P., 2000Probing Photosynthesis: Mechanisms, Regulation, and Adaptation. Taylor & Francis, London, pp. 51–69.

Komatsu, S., Kojima, K., Suzuki, K., Ozaki, K., Higo, K., 2004. Rice proteome database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. Nucl. Acids Res. 32, D388–D392.

Kopriva, S., Bauwe, H., 1995. Serine hydroxymethyltransferase from *Solanum tuberosum*. Plant Physiol. 107, 271–272.

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

Leone, A., Piro, G., Leucci, M.R., Zacheo, G., Dalessandro, G., 2000. Membrane-cell wall-associated heat shock proteins in two genotypes of barley seedlings. Plant Biosyst. 134, 171–178.

Lin, S., Chang, M., Tsai, Y., Lur, H., 2005. Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. Proteomics 5, 2140–2156.

Lonosky, P., Zhang, X., Honavar, V., Dobbs, D., Fu, A., Rodermel, S., 2004. A proteomic analysis of maize chloroplast biogenesis. Plant Physiol. 134, 560–574.

Makino, A., Mae, T., Ohira, K., 1985. Photosynthesis and ribulose-1,5-bisphosphate carboxylase/oxygenase in rice leaves from emergence through senescence. Planta 166, 414–420.

Mason, H.S., Guerrero, F.D., Boyer, J.S., Mullet, J.E., 1988. Proteins homologous to leaf glycoproteins are abundant in stems of darkgrown soybean seedlings. Analysis of proteins and cDNAs. Plant Mol. Biol. 11, 845–856.

Mathesius, U., Keijzers, G., Natera, S.H.A., Winman, J.J., Djordjevic, M.A., Rolfe, B.G., 2001. Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. Proteomics 1, 1424–1440.

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

Molloy, M.P., Herbert, B.R., Walsh, B.J., Tyler, M.I., Sanchez, J.C., Hochstrasser, D.F., Williams, K.L., Gooley, A.A., 1998. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis 19, 837–844.

Mooney, B.P., Thelen, J.J., 2004. High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. Phytochemistry 65, 1733–1744.

Natarajan, S., Xu, C., Caperna, T.J., Garrett, W.M., 2005. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Anal. Biochem. 342, 214–220.

Newsholme, S.J., Maleeft, B.F., Steiner, S., Anderson, N.L., Schwartz, L.W., 2000. Two-dimensional electrophoresis of liver proteins: characterization of a drug-induced hepatomegaly in rats. Electrophoresis 21, 2122–2128.

Newton, R.P., Brenton, A.G., Smith, C.J., Dudley, E., 2004. Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. Phytochemistry 65, 1449–1485.

Pandey, A., Mann, M., 2000. Proteomics to study genes and genomes. Nature 405, 837–846.

Perkins, D.N., Pappinl, D.J.C., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Porubleva, L., Velden, K.V., Kothari, S., Oliver, D.J., Chitnis, P.R., 2001. The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. Electrophoresis 22, 1724–1738.

Santoni, V., Molloy, M., Rabilloud, T., 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis 21, 1054–1070.

Sarnighausen, E., Wurtz, V., Heintz, D., Dorsselaer, A.V., Resk, R., 2004. Mapping of the *Physcomitrella patens* proteome. Phytochemistry 65, 1589–1607.

Schiltz, S., Gallardo, K., Huart, M., Negroni, L., Sommerer, N., Burstin, J., 2004. Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. Plant Physiol. 135, 2241–2260.

Stevens, T.J., Arkin, I.T., 2000. Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins 39, 417–420.

Taiz, L., Zeiger, E., 2002. Photosynthesis: the light reactions. In: Plant physiology. Sinauer Associates, pp. 134–135 (Chapter 7).

Wallin, E., von Heijne, G., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 7, 1029–1038.

Wang, Y., Sun, J., Chitnis, P.R., 2000. Proteomic study of the peripheral proteins from thylakoid membranes of the cyanobacterium *Synechocystis* sp. PCC 6803. Electrophoresis 21, 1746–1754.

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

Wilson, K.A., McManus, M.T., Gordon, M.E., Jordan, T.W., 2002. The proteomics of senescence in leaves of white clover, *Trifolium repens* (L.). Proteomics 2, 1114–1122.

Yamaguchi, K., von Knoblauch, K., Subramanian, A.R., 2000. The plastid ribosomal proteins: identification of all the proteins in the 30S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. 275, 28455–28465.

Yamaguchi, K., Prieto, S., Beligni, M., Haynes, P.A., McDonald, W.H., Yates, J.R., Mayfield, S., 2002. Proteomic characterization of the small subunit of *Chlamydomonas reinhardtii* chloroplast ribosome: identification of a novel S1 domain-containing protein and unusually large orthologs of bacterial S2, S3, and S5. Plant Cell 14, 2957–2976.