Available online at www.sciencedirect.com
SCIENCE DIRECT®
PHYTOCHEMISTRY
Phytochemistry 65 (2004) 1589–1607
www.elsevier.com/locate/phytochem

Mapping of the *Physcomitrella patens* proteome

Eric Sarnighausen [a,*,1], Virginie Wurtz [b,1], Dimitri Heintz [a], Alain Van Dorsselaer [b], Ralf Reski [a]

[a] *Plant Biotechnology, University of Freiburg, Schaenzlestr. 1, 79104 Freiburg, Germany*
[b] *Laboratoire de spectrométrie de masse Bio-Organique, CNRS-UMR 7509/ULP, 25 rue Becquerel, 67087 Strasbourg cedex 2, France*

Received 14 January 2004; received in revised form 2 April 2004
Available online 7 June 2004

Abstract

The moss *Physcomitrella patens* is unique among land plants due to the high rate of homologous recombination in its nuclear DNA. The feasibility of gene targeting makes Physcomitrella an unrivalled model organism in the field of plant functional genomics. To further extend the potentialities of this seed-less plant we aimed at exploring the *P. patens* proteome. Experimental conditions had to be adopted to meet the special requirements connected to the investigations of this moss. Here we describe the identification of 306 proteins from the protonema of Physcomitrella. Proteins were separated by two dimensional electrophoresis, excised form the gel and analysed by means of mass spectrometry. This reference map will lay the basis for further profound studies in the field of Physcomitrella proteomics.

Keywords: *Physcomitrella patens*; Funariaceae; Moss; Functional genomics; Two-dimensional electrophoresis; Mass spectrometry; MALDI-TOF-MS; Nano LC-MS-MS

1. Introduction

The outset of the post genomic era has been initiated by the rather disappointing insight, that gene function cannot be directly discerned from the sequence of the DNA and the tremendous amount of sequence information provided by different genome initiatives has left scientists with the task of assigning function to putative genes (Holtorf et al., 2002). The reverse genetics approach takes its place at the core of functional genomics, i.e. instead of having a function and identifying the responsible gene, a specific gene is manipulated in a way that mirrors its function by modifying the phenotype of the organism under study (Berg, 1991; Tissier and Bourgeois, 2001). Elegant as this approach might be it is likewise rather challenging in both of its requirements, the targeted manipulation of the gene of interest, as well as the precise characterization of the specific influence on the phenotype (Bouche and Bouchez, 2001; Levanon et al., 2003). Although gene function can be targeted temporarily (*knocked down*) by the use of ribozyme, antisense oligonucleotides, aptamers or interfering RNA (RNAi) (Heermeier et al., 2002), the unsurpassed – as permanent and highly efficient – way to the targeted influence of gene function is offered by homologous recombination (Struhl, 1983; Bronson and Smithies, 1994; Kowalczykowski et al., 1994). This process, that involves an exchange of DNA fragments between two molecules at sites of (nearly) identical nucleotide sequence, is applicable as a tool for gene targeting in bacteria, yeast and mammalian embryonic stem cells. It could only recently be demonstrated that in the moss *Physcomitrella patens* foreign DNA sequences are integrated into the nuclear genome

* Corresponding author. Tel.: +49-761-203-6943/2814; fax: +49-761-203-6945.
E-mail address: Eric.Sarnighausen@biologie.uni-freiburg.de (E. Sarnighausen).
[1] These authors contributed equally to this work.

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.04.028

preferentially via homologous recombination at a rate comparable to yeast cells (Schaefer and Zrÿd, 1997). To date, Physcomitrella is the only land plant that has been shown to allow for this direct method of gene targeting and therefore serves as a model organism for plant functional genomics (Reski, 2003). As a model system, Physcomitrella offers additional advantages. The early developmental stage, the filamentous moss protonema, can be grown in axenic liquid culture in a highly standardized manner either in flask culture or in bioreactors of up to 10 litres (Hohe and Reski, 2002). As mosses in general display a heterophasic and heteromorphic alternation of generations with the haploid phase being dominant, complicated backcrossing in order to obtain homozygote plants is unnecessary and genotypic variations can directly affect the phenotype (Reski, 1998). Although the complete genome of *Physcomitrella patens* has not been sequenced yet, a clustered library from 110,087 expressed sequence tags (ESTs) comprising at least 95% of all *P. patens* genes – the number of which is estimated to be around 25,000 – was established by our group in cooperation with an industrial partner. Profound analysis of the database revealed that approximately 50% of the Physcomitrella protein coding genes can be matched to homologues within the Arabidopsis genome. The function of around 11,600 Physcomitrella genes still remains unrevealed, as no currently known homologues in plants, animals, fungi or bacteria can be found (Rensing et al., 2002). Nishiyama et al. (2002) reported the sequencing of 85,191 ESTs from *P. patens*, that, together with publicly available *P. patens* ESTs, could be organized into 15,883 putative transcripts. Efforts are underway to establish saturated Physcomitrella mutant libraries (Nishiyama et al., 2000; Egener et al., 2002). Out of 16,203 mutant plants already analysed by our group, 16.2% display a deviating phenotype as compared to the wild type – be it a more or less conspicuous morphological or developmental alteration or a partial loss of autotrophy that can be compensated for by an additional supply of nutrients (Egener et al., 2002). This number of obvious phenotypical aberrations observed in Physcomitrella mutants is rather high as, for example, compared to a rate of only 2% in case of *Arabidopsis thaliana* knockout mutants (Bouche and Bouchez, 2001). However, the fact that more than 80% of the mutations generated do not seem to have any impact on the phenotype of the plants clearly demonstrates that a mere visual evaluation is far from being sufficient if gene function is to be identified by reverse genetics. And even in cases where a mutation does affect the phenotype obviously, might it still be rather difficult to derive a gene function from the frequently unspecific alterations observed (as, for example, a retardation of growth). While it appears plausible that the complex organism might still be able to tolerate and thereby mask a mutation by some sort of phenotypic compensation, it is rather unlikely that the genotypic disorder imposed on the plant might not at all be reflected by physiological and/or biochemical imbalances (Raamsdonk et al., 2001). Only in a limited number of cases (if ever) will the knockout of a gene simply result in the disappearance of the corresponding gene product. As proteins in a living organisms cannot be viewed as ''lone fighters'', its loss will rather evoke imbalances within functional protein networks possibly culminating in an avalanche effect (Henning and Beste, 2002). It is evidently the induced dysfunction of a complex network that finally governs the phenotype – which will be regarded as non-aberrative if functional compensation within the system is accomplished. A complete exploitation of the rich potential of *P. patens* as a tool for functional genomics will therefore demand for additional appropriate screening procedures in order to define gene function in regulatory complexes and to allow the revelation of the so-called ''silent mutations'' by analysing cellular and higher order functions (Sperling, 2001).

As a part of a Physcomitrella proteomics project we are currently mapping the *P. patens* proteome using combinations of different protein extraction methods, electrophoretic as well as chromatographic separation techniques and mass spectrometric analyses. Here we present our first results towards the identification of the Physcomitrella proteome by means of 2D protein gel electrophoresis and two different mass spectrometry methods.

2. Results and discussion

Here we present for the first time a two-dimensional protein map of the moss *P. patens*. Moss protonema grown in a bioreactor was extracted using acetone/TCA precipitation and the total proteins were separated by isoelectric focusing in the first dimension and by SDS–PAGE in the second dimension. To date, proteomics still relies heavily on 2D electrophoresis. Though frequently denoted as outdated (Gygi et al., 2000, Hille et al., 2001), alternative sorting procedures that have been introduced could so far not establish a position comparable to the 2D method: major advantages of this approach are the high information content obtained during the experiments and the high resolution capability (Fey and Larsen, 2001; Beranova-Giorgianne, 2003). The technique that had been introduced almost 30 years ago could be facilitated enormously with respect to both handling and reproducibility by the introduction of immobilized pH gradient (IPG) in ready cast gels as a first dimension, which opened the path to proteomics even to scientists rather inexperienced in the field (Görg et al., 1988).

If, however, proteins are not only to be visually analysed after two-dimensional gel electrophoresis but their identification by means of mass spectrometry is desired, special care has to be taken in order to avoid methods that would render analysis and identification of the proteins difficult or even impossible. This includes unwanted and/or uncontrollable modifications of the proteins during extraction (e.g. oxidation by endogenous or applied phenolic compounds, carbamylation) (Hari, 1981; McCarthy et al., 2003), electrophoresis (covalent modification by free acrylamide radicals) (Chiari et al., 1990) and detection (covalent modification by aldehydes) (Shevchenko et al., 1996). Phenolic extraction that has been reported to be effective in case of plant tissues including bryophytes (Hurkman and Tanaka, 1986; Hellwege et al., 1996) is therefore definitely not recommended. Classical high sensitive silver staining methods that include treatment of the gels with glutaraldehyde and formaldehyde cannot be used either. Protein extraction was therefore performed using acetone/TCA precipitation as described for tissues from seed plants. The method instantly inactivates plant proteases and phenoloxidases thereby preventing protein degradation and polymerization of phenolic compounds and at the same times clears the extract from interfering substances (Granier, 1988). Apart from yielding to high quality 2D electropherograms, the TCA/acetone extraction could be shown not to interfere with mass spectrometric analysis and identification of proteins (Porubleva et al., 2001, Koller et al., 2002). Two dimensional gels were finally stained with colloidal coomassie if proteins were to be extracted for mass spectrometry. As it turned out, analytical gels loaded with less amounts of total protein and subjected to sensitive silver staining displayed patterns that differed strongly from coomassie stained gels. This is mainly due to a large number of protein spots that are not stainable with silver and therefore escaped detection (not shown).

Out of 790 spots that could be resolved and displayed on a Coomassie stained gel (Fig. 1(a)), 267 spots of

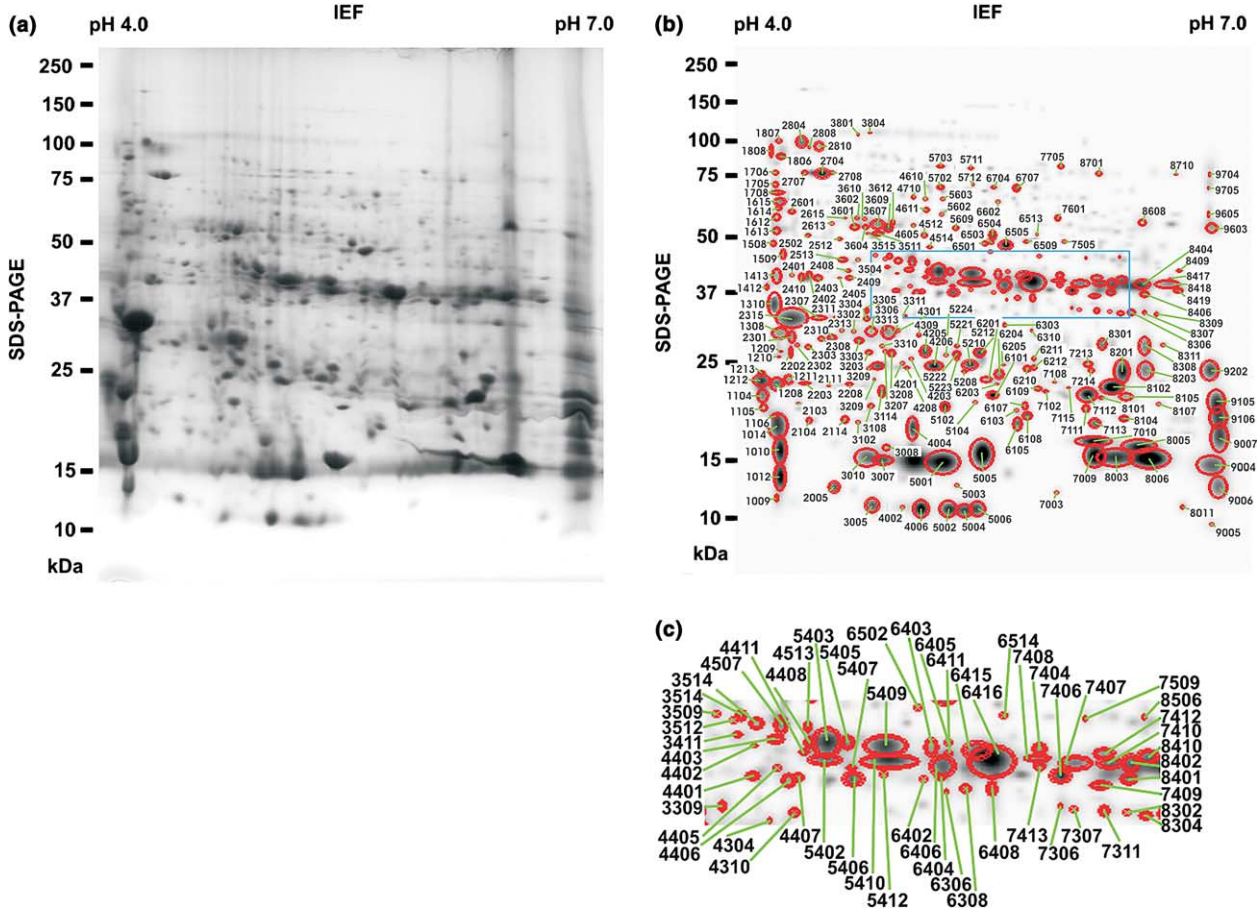

Fig. 1. Two-dimensional electropherogram of total proteins extracted form the protonema of *P. patens*. (a) 2D-PAGE gel stained with Coomassie Brillant Blue. (b) Computer generated image (gaussian image) of the same gel displaying protein spots selected for analysis by mass spectrometry. Standard spot numbers (SSPs) were assigned to each spot by the analysis software PDQuest. (c) In order to reduce the complexity of the images, SSPs of the protein spots boxed in (b) are displayed separately.

moderate to high staining intensity were selected, cut from the gel, digested with trypsin and analysed by mass spectrometry (MALDI-TOF-MS and Nano-LC-MS-MS). 208 of these 2D spots yielded interpretable mass spectra (Fig. 1(b)). The distribution of the distinct spot quantities (as determined by the PDQuest software) contributes to the total stained area investigated in an exponential manner (Fig. 2). It is obvious that of the 208 spots, a total number of 181 spots of low (153 counts/[Image Units]^2) to medium (2144 counts/IU^2) quantity build 50% of the area of investigation, whereas the remaining 50% are covered by 27 high quantity spots (up to 6878 counts/IU^2).

The spots analysed are of molecular masses between 11,400 and 100,900 where the size range from 20,000 to 45,000 is represented by 130 spots out of 208 (Fig. 3).

From 208 spots 306 distinct proteins could be identified. In a complex protein pattern (as is unavoidable to be found with total protein extracts), the limits of 2D electrophoresis are evident in cases of such spot overlaps, i.e. when two different proteins share close similarity with respect to both mass and isoelectric point and therefore comigrate to the same position on the gel. In the case of the Physcomitrella electropherogram, 29% of all spots analysed contained two different proteins, while 17% of the spots contained three and 0.5% (one spot) even four different proteins. A clear distinction and detection of multiple proteins in these spot samples is, however, possible with the use of a Nano-LC-MS-MS mass spectrometer. Although therefore in our studies, comigration of proteins does not really raise a problem, the usefulness of 2D electrophoresis for differential proteomics (i.e. the detection of deviations in protein patterns of two or more different samples) will definitely benefit from an increase in resolution in the case of Physcomitrella protonema. This can be accomplished by the use of narrow pH gels in the first dimension (Wildgruber et al., 2000) or a prefractionation of the sample prior to electrophoresis (Molloy et al., 1998).

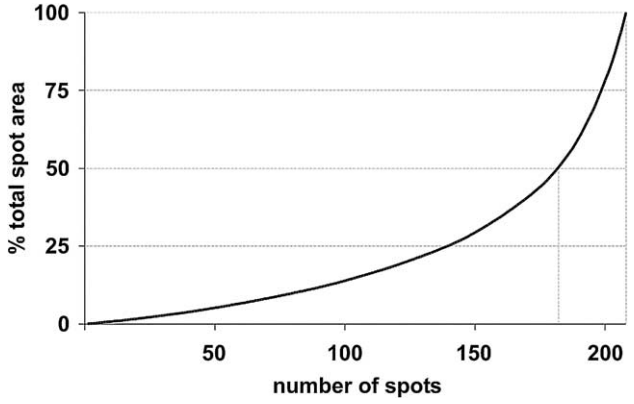

Fig. 2. Distribution of quantities among the spots selected for analysis. The abscissa displays the spots analysed arranged in order of their quantities from low (153) to high (6878) values. The graph represents the integration of single spot quantities (total spot area).

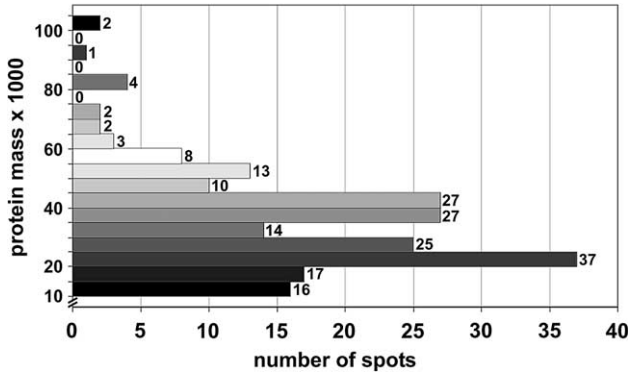

Fig. 3. Distribution of molecular masses among the spots selected for analysis. Spots are grouped in classes each covering 5000 Da.

Protein identification was performed by both peptide mass fingerprinting on a MALDI-TOF mass spectrometer exploiting the non redundant database SWISS-PROT (http://www.expasy.ch/sprot) using MASCOT (Perkins et al., 1999) search program as well as MS-MS sequencing with a Q-TOF mass spectrometer, where correlation of uninterpreted tandem mass spectra were correlated to entries in SWISS-PROT. While the strength of MALDI-TOF-MS is its suitability for high throughput analysis, the identification procedure in the ideal case requires for a database comprising the complete (specifically digested) proteome of the organism under study for comparison (Langen and Berndt, 2001). In the case of Physcomitrella, large parts of the coding region of the genome are available as EST data. While this does – in contrast to the use of higher eukaryotic genomic libraries – render the prediction of splicing events unnecessary, each EST has to be translated into protein from each of the six possible reading frames. This gives rises to an inflation of the deduced protein database with 5/6th of ''nonsense information'' that can act as a rich source of false positive matches. At the same time, deduced protein databases do not take into account to co- and post-translational modifications that induce alterations of peptide masses.

It is therefore not too surprising that a much more reliable identification of the proteins can be obtained by the – though more time consuming and demanding – MS-MS sequencing approach. Of the 306 proteins described, 26% (80) could be identified by MALDI-TOF MS. In most cases, (18% [56]) MALDI-TOF-MS results were confirmed by MS-MS spectra. In 17 distinct Coomassie stainable spots (6%), however, proteins were

identified by MALDI-TOF-MS, the presence of which could not be detected via MS-MS search. In 2% of the cases, proteins were identified by MALDI-TOF-MS while MS-MS spectra failed completely.

Identified proteins were grouped according to their function as described by the EU Arabidopsis Genome Project (1998) (Fig. 4). Not surprisingly for this approach, the majority of proteins detected could be assigned to housekeeping functions like energy (functional catalogue 2) and metabolism (functional catalogue 1). It is, however, intriguing that even among the relatively highly abundant proteins that have been analysed so far, 46 proteins in 40 spots correspond to Physcomitrella ESTs that do either not yield a significant match in a BLAST search or correspond to genes encoding hypothetical, putative or unknown proteins (Table 1). Multiple spots obviously corresponding to one gene product are commonly found on a 2D electropherogram (Fountoulakis et al., 2001). There are three major reasons that are responsible for this phenomenon. These spots might not be the products of a single gene but are derived from different members of a multigene family. The distinct biophysical properties that govern their migration behaviour in the two dimensions are therefore due to aberrations in the amino acid sequences. Likewise is it possible that one gene product undergoes different co- and/or posttranslational modifications that affect its mass or/and its charge. It can, however, not be ruled out that the appearance of multiple spots corresponding to one gene product is merely a consequence of the introduction of artificial modifications during the extraction or separation procedure (Gooley and Packer, 1997, Berven et al., 2003). The assignment of such gene products to different expressed sequence tags or even different contigs hints to the existence of a multigene family. Chances to identify and distinguish between the members of this protein family improve with the amount of primary amino acid sequence covered by MS-MS sequencing. Fig. 5 provides an example for proteins from eight distinct spots that were identified as 23 kDa polypeptide of the oxygen evolving complex. The alignment of the de novo sequences with the corresponding translation products of ESTs and contigs clearly demonstrates that the heterogeneity of the proteins is partly due to the fact that they are derived from (at least) three distinct but closely related genes.

The success of this strategy is crucially dependent on using MS-MS data as the query because, in contrast to peptide mass fingerprinting (PMF), which relies on the mass measurement of multiple peptides for statistically valid identification, one or two matching MS-MS spectra are often sufficient. Thus, 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.

Of the 267 spots successfully excised from the gel, a total number of 59 did not yield interpretable mass spectra. It should be noted that most of these spots displayed either low quantity or low spot quality. It remains to be speculated if the proteins were heavily modified or possibly resisted tryptic digestion and/or extraction from the gel pieces.

To date, plant science is highly underrepresented in the field of proteomics. The Swiss 2D PAGE map selection at the ExPASy proteomics server (at http://www.expasy.org/cgi-bin/map1) is dominated by human,

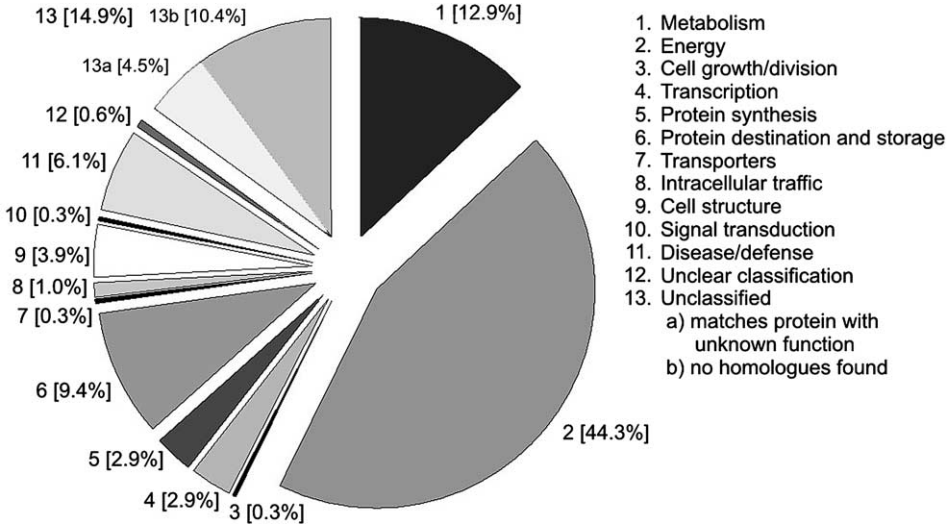

Fig. 4. Pie chart displaying the assignment of proteins identified to functional categories according to EU Arabidopsis Genome Project (1998) (100% = 306 proteins).

Table 1
Proteins from the protonema of *P. patens* identified by mass spectrometry after 2D electrophoresis

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
Functional category 1: metabolism							
3303	28.3	5.18	712	Nano LC-MS-MS	7047325	PPP_1186_C1	3-β-Hydroxysteroid dehydrogenase/ isomerase protein, putative
6514	45.2	6.10	189	Nano LC-MS-MS	7583616	PPP_3143_C1	Acetylornithine transaminase, putative
7108	22.8	6.18	783	Nano LC-MS-MS	18335736	PPP_6841_C1	β-Hydroxyacyl-ACP dehydratase, putative
7115	22.1	6.25	265	Nano LC-MS-MS	18335736	PPP_6841_C1	β-Hydroxyacyl-ACP dehydratase, putative
6513	51.4	6.09	206	Nano LC-MS-MS	14536525	No hits	Biotin carboxylase
4512	51.1	5.51	525	Nano LC-MS-MS	18367049	PPP_7060_C1	Carbamoyl phosphate synthase, putative
4406	36.8	5.47	167	Nano LC-MS-MS	18339112	PPP_1097_C1	Cysteine synthase 1, plastidic
4406	36.8	5.47	167	Nano LC-MS-MS	18356110	PPP_8043_C1	Cysteine synthase, mitochondrial precursor
4406	36.8	5.47	167	MALDI-TOF-MS	[O81155]	nd	Cysteine synthase, chloroplast precursor
9603	53.7	6.97	1245	Nano LC-MS-MS	18333013	PPP_516_C1	Dihydrolipoamide dehydrogenase precursor
6308	35.9	5.99	189	Nano LC-MS-MS	14536601	PPP_sd_136_C1	Enoyl-[acyl-carrier protein] reductase
6308	35.9	5.99	189	MALDI-TOF-MS	18327390	PPP_sd_136_C1	Enoyl-[acyl-carrier protein] reductase
8506	45.0	6.52	529	Nano LC-MS-MS	18342459	PPP_11879_C1/ PPP_8061_C1	Glutamate-1-semialdehyde 2,1-aminomutase, chloroplast precursor
4514	47.6	5.54	290	MALDI-TOF-MS	[P24630]	nd	Glutamate-1-semialdehyde 2,1-aminomutase
2409	40.2	5.14	311	Nano LC-MS-MS	18332800	PPP_19_C1	Glutamine synthetase
8402	39.0	6.48	1159	Nano LC-MS-MS	18332800	PPP_19_C1	Glutamine synthetase
8417	38.8	6.77	222	Nano LC-MS-MS	18332800	PPP_19_C1	Glutamine synthetase
8419	37.5	6.83	664	Nano LC-MS-MS	18332800	PPP_19_C1	Glutamine synthetase
2401	40.7	4.64	187	Nano LC-MS-MS	18333729	PPP_19_C1	Glutamine synthetase
8418	39.8	6.79	1374	Nano LC-MS-MS	18333729	PPP_19_C1	Glutamine synthetase
1413	40.5	4.39	1158	Nano LC-MS-MS	18333924	PPP_19_C1	Glutamine synthetase
7407	39.0	6.32	1368	Nano LC-MS-MS	18333924	PPP_19_C1	Glutamine synthetase
7410	38.9	6.42	1022	Nano LC-MS-MS	18333924	PPP_19_C1	Glutamine synthetase
7413	38.6	6.21	1373	Nano LC-MS-MS	18333924	PPP_19_C1	Glutamine synthetase
8404	38.8	6.64	4861	Nano LC-MS-MS	18333924	PPP_19_C1	Glutamine synthetase
6408	36.0	6.07	168	Nano LC-MS-MS	18371927	BJ203529	Hypothetical protein SAV0519 "conserved hypothetical protein"
2315	32.1	4.63	5355	Nano LC-MS-MS	15390915	PPP_3564_C1	Inorganic pyrophosphatase
2408	41.7	5.13	232	Nano LC-MS-MS	18333268	No hits	Magnesium chelatase subunit of protochlorophyllide reductase
2408	41.7	5.13	232	MALDI-TOF-MS	[P93162]	nd	Magnesium chelatase subunit of protochlorophyllide reductase
6107	20.2	6.03	621	MALDI-TOF-MS	18332406	PPP_3209_C3	Nucleoside diphosphate kinase II
6103	19.8	5.98	309	Nano LC-MS-MS	18341748	PPP_3209_C3	Nucleoside diphosphate kinase II
6103	19.8	5.98	309	MALDI-TOF-MS	18332406	PPP_504_C1	Nucleoside diphosphate kinase II
6105	18.5	5.99	1317	Nano LC-MS-MS	18341748	PPP_3209_C3	Nucleoside diphosphate kinase II
7113	18.6	6.39	849	Nano LC-MS-MS	18333449	PPP_2923_C3	Nucleoside diphosphate kinase II
7113	18.6	6.39	849	MALDI-TOF-MS	18333449	PPP_2923_C3	Nucleoside diphosphate kinase II
4304	32.6	5.42	202	Nano LC-MS-MS	18335703	PPP_1876_C1	Protein F1N21.10 [imported]
2310	28.8	5.05	758	Nano LC-MS-MS	18339207	PPP_sd_94_C1	Ribose-5-phosphate isomerase precursor
2308	28.7	4.99	350	Nano LC-MS-MS	18366345	PPP_sd_94_C1	Ribose-5-phosphate isomerase precursor
6501	48.5	5.82	182	MALDI-TOF-MS	[P49611]	nd	*S*-Adenosylmethionine synthetase 1
6501	48.5	5.82	182	Nano LC-MS-MS	18361272	PPP_7319_C3	*S*-Adenosylmethionine synthetase 1
6503	48.6	5.85	351	Nano LC-MS-MS	18331526	PPP_7319_C3	*S*-Adenosylmethionine synthetase 1
6503	48.6	5.85	351	MALDI-TOF-MS	[P23686]	nd	*S*-Adenosylmethionine synthetase 1
6505	48.1	5.92	1561	Nano LC-MS-MS	18341203	PPP_229_C2	*S*-Adenosylmethionine synthetase 1
6505	48.1	5.92	1561	Nano LC-MS-MS	18351002	PPP_7319_C3	*S*-Adenosylmethionine synthetase 1
6509	49.0	6.03	408	Nano LC-MS-MS	18341203	PPP_229_C2	*S*-Adenosylmethionine synthetase 1
6509	49.0	6.03	408	MALDI-TOF-MS	[P23686]	nd	*S*-Adenosylmethionine synthetase 1

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
4406	36.8	5.47	167	Nano LC-MS-MS	7148079	PPP_sd_164_C1	*S*-Malonyltransferase [acyl-carrier protein]
6408	36.0	6.07	168	Nano LC-MS-MS	18339778	PPP_272_C1	SOR1 from the fungus Cercospora nicotianae protein, putative
3612	55.8	5.35	548	Nano LC-MS-MS	18357422	PPP_7529_C1	UDP-glucose pyrophosphorylase
Functional category 2: energy							
8506	45.0	6.52	529	Nano LC-MS-MS	18333687	PPP_2459_C1	12-Oxophytodienoate reductase, similar to
8201	24.0	6.53	2366	MALDI-TOF-MS	18340444	PPP_1619_C1	23 kDa polypeptide of oxygen-evolving comlex (OEC)
8201	24.0	6.53	2366	Nano LC-MS-MS	18340444	PPP_1619_C1	23 kDa polypeptide of OEC
4206	24.6	5.56	2144	Nano LC-MS-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
6204	23.6	5.89	953	Nano LC-MS-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
7108	22.8	6.18	783	MALDI-TOF-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
7108	22.8	6.18	783	Nano LC-MS-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
7115	22.1	6.25	265	Nano LC-MS-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
8102	22.2	6.48	3978	MALDI-TOF-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
8102	22.2	6.48	3978	Nano LC-MS-MS	18341623	PPP_2787_C1	23 kDa polypeptide of OEC
9202	24.1	6.96	4888	Nano LC-MS-MS	18335637	PPP_678_C1	23 kDa polypeptide of OEC
6205	24.7	5.91	630	Nano LC-MS-MS	18336607	PPP_678_C1	23 kDa polypeptide of OEC
2315	32.1	4.63	5355	MALDI-TOF-MS	18340210	PPP_655_C3	33kDa oxygen evolving protein of photosystem II, putative
2315	32.1	4.63	5355	Nano LC-MS-MS	18340210	PPP_655_C3	33kDa oxygen evolving protein of photosystem II, putative
5402	39.4	5.58	649	Nano LC-MS-MS	18332642	AW145742	Aldolase, putative
6411	40.2	5.93	736	Nano LC-MS-MS	18337795	AW145742	Fructose-bisphosphate aldolase, putative
5402	39.4	5.58	649	Nano LC-MS-MS	[Q40677]	nd	Fructose-bisphosphate aldolase
6406	38.6	5.92	443	Nano LC-MS-MS	18335784	PPP_3480_C1	Aldolase, plastidic
8418	39.8	6.79	1374	Nano LC-MS-MS	18357689	PPP_436_C1	Fructose-bisphosphate aldolase
5402	39.4	5.58	649	Nano LC-MS-MS	7207467	PPP_sd_105_C1	Aldolase-like protein, latex plastidic
6406	38.6	5.92	443	Nano LC-MS-MS	7207467	PPP_sd_105_C1	Aldolase-like protein, latex plastidic
6411	40.2	5.93	736	Nano LC-MS-MS	18330041	PPP_sd_105_C1	Aldolase, putative
3204	23.1	5.26	298	MALDI-TOF-MS	7066995	PPP_2006_C1	ATP synthase, H^+-transporting putative
3204	23.1	5.26	298	Nano LC-MS-MS	18334058	PPP_2006_C1	ATP synthase, H^+-transporting putative
3604	53.9	5.22	172	Nano LC-MS-MS	[P06284]	nd	ATP synthase β chain
3601	54.1	5.16	302	MALDI-TOF-MS	[Q31794]	nd	ATP synthase β chain
3601	54.1	5.16	302	Nano LC-MS-MS	[Q31794]	nd	ATP synthase β chain
3604	53.9	5.22	172	Nano LC-MS-MS	[Q31794]	nd	ATP synthase β chain
6205	24.7	5.91	630	Nano LC-MS-MS	[Q43432]	nd	ATP synthase, vacuolar subunit B isoform 1
3607	55.2	5.28	1543	Nano LC-MS-MS	18350686	PPP_7338_C1	ATPase β subunit
3609	53.6	5.33	681	Nano LC-MS-MS	18350686	PPP_7338_C1	ATPase β subunit
8301	27.7	6.43	1055	Nano LC-MS-MS	18326953	PPP_3174_C1	Carbonate dehydratase (EC 4.2.1.1), probable
6310	30.0	6.06	330	Nano LC-MS-MS	18348698	PPP_3174_C1	Carbonate dehydratase (EC 4.2.1.1), probable
3313	30.4	5.31	511	Nano LC-MS-MS	18327334	PPP_8051_C1	Carbonic anhydrase 1
3313	30.4	5.31	511	Nano LC-MS-MS	18336080	PPP_3459_C1	Carbonic anhydrase
3302	29.8	5.16	516	Nano LC-MS-MS	18360868	PPP_8051_C1	Carbonic anhydrase 1
2301	28.9	4.64	288	Nano LC-MS-MS	7067286	PPP_9428_C1	Chlorophyll *a*/*b*-binding protein precursor
7112	21.4	6.35	2205	Nano LC-MS-MS	16137492	PPP_1943_C1	Cytochrome B6-F complex iron-sulfur subunit
5609	53.7	5.67	323	Nano LC-MS-MS	18329259	PPP_7388_C1	Enolase, α
5609	53.7	5.67	323	Nano LC-MS-MS	18332690	PPP_3959_C1	Enolase, putative
6503	48.6	5.85	351	Nano LC-MS-MS	18338306	PPP_608_C2	Enolase (2-phospho-D-glycerate hydroylase)
6504	50.8	5.86	783	Nano LC-MS-MS	18338306	PPP_608_C2	Enolase (2-phospho-D-glycerate hydroylase)

(*continued on next page*)

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
6504	50.8	5.86	783	MALDI-TOF-MS	18341535	PPP_608_C2	Enolase (2-phospho-D-glycerate hydroylase)
6503	48.6	5.85	351	Nano LC-MS-MS	18361876	PPP_7498_C3	Enolase
4512	51.1	5.51	525	Nano LC-MS-MS	18365747	PPP_7498_C3	Enolase
1212	23.0	4.00	592	Nano LC-MS-MS	2065364	PPP_sd_185_C1	Ferredoxin, chloroplast
7311	33.5	6.40	523	MALDI-TOF-MS	18337946	PPP_166_C2	Ferrodoxin NADP oxidoreductase
7311	33.5	6.40	523	Nano LC-MS-MS	18335779	PPP_166_C2	Ferrodoxin NADP oxidoreductase
8306	33.1	6.59	929	Nano LC-MS-MS	18335779	PPP_166_C2	Ferrodoxin NADP oxidoreductase
8309	32.9	6.72	337	Nano LC-MS-MS	18335779	PPP_166_C2	Ferrodoxin NADP oxidoreductase
9105	20.7	6.99	4518	Nano LC-MS-MS	18362604	PPP_2786_C2	Germin-like protein
9105	20.7	6.99	4518	MALDI-TOF-MS	15261154	PPP_429_C1	Germin protein, putative
6109	22.1	6.09	676	Nano LC-MS-MS	18364211	PPP_429_C1	Germin protein, putative
8406	36.6	6.65	224	Nano LC-MS-MS	18336775	PPP_29_C1	Malate dehydrogenase
6404	37.4	5.91	638	Nano LC-MS-MS	18340386	PPP_2224_C1	Malate dehydrogenase (EC 1.1.1.37), nodule-enhanced
6404	37.4	5.91	638	Nano LC-MS-MS	[Q42686]	nd	Malate dehydrogenase, mitochondrial precursor
6404	37.4	5.91	638	Nano LC-MS-MS	18371317	PPP_7351_C1	Malate dehydrogenase NAD-MDH, chloroplast
1412	38.3	4.15	429	Nano LC-MS-MS	18349266	PPP_8920_C1	NADH-ubiquinone oxidoreductase, putative
5223	25.6	5.68	407	Nano LC-MS-MS	18335711	PPP_5639_C1	Oxygen-evolving complex 25.6 kD protein, chloroplast precursor,
1105	20.1	4.08	544	Nano LC-MS-MS	7147909	PPP_2103_C1	Oxygen-evolving enhancer protein 3, precursor-like protein
1105	20.1	4.08	544	Nano LC-MS-MS	15327293	PPP_2250_C1	Oxygen-evolving enhancer protein 3, precursor-like protein
3509	45.5	5.27	237	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
3512	44.6	5.31	254	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
3514	45.2	5.34	485	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
4501	44.1	5.38	301	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
4507	44.3	5.45	1294	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
4513	43.7	5.53	383	Nano LC-MS-MS	14597787	PPP_sd_87_C1	Phosphoglycerate kinase, chloroplast precursor
5405	41.5	5.64	1291	Nano LC-MS-MS	7286982	PPP_347_C2	Phosphoribulokinase, chloroplast precursor
3005	12.9	5.25	1641	Nano LC-MS-MS	[P56301]	nd	Photosystem I iron-sulfur center
4006	12.1	5.50	3581	MALDI-TOF-MS	[P56301]	nd	Photosystem I iron-sulfur center
4006	12.1	5.50	3581	Nano LC-MS-MS	[P56301]	nd	Photosystem I iron-sulfur center
5002	11.9	5.63	4506	Nano LC-MS-MS	[P56301]	nd	Photosystem I iron-sulfur center
5004	11.4	5.71	1150	Nano LC-MS-MS	[P56301]	nd	Photosystem I iron-sulfur center
5006	12.1	5.77	715	MALDI-TOF-MS	[P56301]	nd	Photosystem I iron-sulfur center
4002	12.7	5.40	326	Nano LC-MS-MS	[Q9TKV9]	nd	Photosystem I iron-sulfur center
5002	11.9	5.63	4506	MALDI-TOF-MS	[Q9TKV9]	nd	Photosystem I iron-sulfur center
7111	20.0	6.35	747	MALDI-TOF-MS	7047306	PPP_11507_C1	Photosystem I subunit II
8101	21.0	6.43	543	MALDI-TOF-MS	15730321	PPP_11507_C1	Photosystem I subunit II
8104	19.0	6.55	555	MALDI-TOF-MS	15730321	PPP_11507_C1	Photosystem I subunit II
9106	19.1	7.00	6627	MALDI-TOF-MS	15730321	PPP_11507_C1	Photosystem I subunit II
1104	21.3	4.04	340	Nano LC-MS-MS	18337985	PPP_11507_C1	Photosystem I subunit II
1208	22.5	4.38	1088	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
1212	23.0	4.00	592	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
3114	21.7	5.30	1161	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
7111	20.0	6.35	747	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
8101	21.0	6.43	543	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
8104	19.0	6.55	555	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
9106	19.1	7.00	6627	Nano LC-MS-MS	18332627	PPP_2303_C2	Photosystem I subunit II
5005	16.0	5.80	3043	Nano LC-MS-MS	18325241	PPP_1539_C1	Photosystem I reaction center subunit IV

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
3008	16.5	5.32	579	Nano LC-MS-MS	18337895	PPP_1539_C1	Photosystem I reaction center subunit IV
9004	14.8	6.97	6878	Nano LC-MS-MS	18369352	PPP_1615_C1	Photosystem I reaction center subunit IV
1012	14.2	4.50	4478	Nano LC-MS-MS	7584262	PPP_6668_C1	Plastocyanin precursor
2402	41.0	4.85	390	Nano LC-MS-MS	18341651	PPP_6487_C1	Quinone oxidoreductase-like protein
5410	39.2	5.76	2233	Nano LC-MS-MS	584290	PPP_1718_C5	RuBisCO activase
3514	45.2	5.34	485	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
4403	42.0	5.43	399	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
4710	67.0	5.46	228	MALDI-TOF-MS	7067400	PPP_1718_C5	RuBisCO activase
5403	41.6	5.58	2497	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
5410	39.2	5.76	2233	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
6415	40.5	6.02	6150	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
6416	39.3	6.07	525	Nano LC-MS-MS	7067400	PPP_1718_C5	RuBisCO activase
3411	42.5	5.33	369	Nano LC-MS-MS	7147876	PPP_1718_C5	RuBisCO activase
5402	39.4	5.58	649	Nano LC-MS-MS	7147876	PPP_1718_C5	RuBisCO activase
5407	38.3	5.65	386	Nano LC-MS-MS	7147876	PPP_1718_C5	RuBisCO activase
7408	39.6	6.20	158	Nano LC-MS-MS	7147876	PPP_1718_C5	RuBisCO activase
8401	37.2	6.48	382	Nano LC-MS-MS	7147876	PPP_1718_C5	RuBisCO activase
8410	39.7	6.54	1070	Nano LC-MS-MS	7207628	PPP_1718_C5	RuBisCO activase
6416	39.3	6.07	525	Nano LC-MS-MS	7584290	PPP_1718_C5	RuBisCO activase
8401	37.2	6.48	382	Nano LC-MS-MS	7584290	PPP_1718_C5	RuBisCO activase
3514	45.2	5.34	485	Nano LC-MS-MS	16137454	PPP_1718_C5	RuBisCO activase
4408	41.6	5.53	307	Nano LC-MS-MS	16137454	PPP_1718_C5	RuBisCO activase
5402	39.4	5.58	649	Nano LC-MS-MS	16137454	PPP_1718_C5	RuBisCO activase
4403	42.0	5.43	399	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4403	42.0	5.43	399	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4408	41.6	5.53	307	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4408	41.6	5.53	307	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4411	40.3	5.52	538	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4710	67.0	5.46	228	MALDI-TOF-MS	16137547	PPP_1718_C5	RuBisCO activase
5403	41.6	5.58	2497	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
5409	41.3	5.75	3692	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
6403	40.9	5.88	428	Nano LC-MS-MS	16137547	PPP_1718_C5	RuBisCO activase
4403	42.0	5.43	399	Nano LC-MS-MS	18331604	PPP_1718_C5	RuBisCO activase
5402	39.4	5.58	649	Nano LC-MS-MS	18331604	PPP_1718_C5	RuBisCO activase
7412	40.2	6.40	665	Nano LC-MS-MS	18331604	PPP_1718_C5	RuBisCO activase
8401	37.2	6.48	382	Nano LC-MS-MS	18331604	PPP_1718_C5	RuBisCO activase
6415	40.5	6.02	6150	Nano LC-MS-MS	18334238	PPP_1718_C5	RuBisCO activase
4408	41.6	5.53	307	Nano LC-MS-MS	[O49074]	PPP_1718_C5	RuBisCO activase
5402	39.4	5.58	649	Nano LC-MS-MS	[O98997]	PPP_1718_C5	RuBisCO activase
7408	39.6	6.20	158	Nano LC-MS-MS	[O98997]	PPP_1718_C5	RuBisCO activase
4403	42.0	5.43	399	MALDI-TOF-MS	[P93431]	PPP_1718_C5	RuBisCO activase
4710	67.0	5.46	228	Nano LC-MS-MS	[P93431]	PPP_1718_C5	RuBisCO activase
7412	40.2	6.40	665	MALDI-TOF-MS	[Q40281]	PPP_1718_C5	RuBisCO activase
6513	51.4	6.09	206	Nano LC-MS-MS	7047489	PPP_1718_C5	RuBisCO activase
3514	45.2	5.34	485	Nano LC-MS-MS	18337005	PPP_2028_C1	RuBisCO activase
3514	45.2	5.34	485	Nano LC-MS-MS	18338667	PPP_2028_C1	RuBisCO activase
1612	58.0	4.38	665	Nano LC-MS-MS	7209742	No hits found	Rubisco large subunit
1612	58.0	4.38	665	Nano LC-MS-MS	[P28422]	nd	Rubisco large subunit
7408	39.6	6.20	158	MALDI-TOF-MS	18335828	BG361035	RuBisCo small subunit
5001	14.9	5.61	6871	MALDI-TOF-MS	16137671	BI894535	Rubsico small subunit
3203	26.6	5.24	462	MALDI-TOF-MS	18333860	BJ165876	Rubisco small subunit
7111	20.0	6.35	747	MALDI-TOF-MS	18333860	BJ165876	Rubisco small subunit
7408	39.6	6.20	158	Nano LC-MS-MS	[O98997]	nd	RuBisCo small subunit
8003	15.3	6.50	3251	Nano LC-MS-MS	[P19308]	nd	Rubisco small subunit
7111	20.0	6.35	747	MALDI-TOF-MS	7067301	PPP_sd_8_CB4	Rubisco small subunit
8006	15.1	6.68	5793	MALDI-TOF-MS	7067301	PPP_sd_8_CB4	Rubisco small subunit
8608	55.7	6.64	2507	MALDI-TOF-MS	15590642	PPP_sd_8_CB4	Rubisco small subunit
3007	15.0	5.31	1335	Nano LC-MS-MS	16137513	PPP_sd_8_CB4	Rubisco small subunit
7111	20.0	6.35	747	Nano LC-MS-MS	16137513	PPP_sd_8_CB4	Rubisco small subunit
8307	33.2	6.65	858	Nano LC-MS-MS	16137513	PPP_sd_8_CB4	Rubisco small subunit

(*continued on next page*)

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
3008	16.5	5.32	579	Nano LC-MS-MS	16137518	PPP_sd_8_CB4	Rubisco small subunit
7306	34.0	6.27	292	MALDI-TOF-MS	16137527	PPP_sd_8_CB4	RuBisCo small subunit
5224	26.1	5.62	240	Nano LC-MS-MS	16137530	PPP_sd_8_CB4	RuBisCo small subunit
3010	15.4	5.22	2504	MALDI-TOF-MS	16137614	PPP_sd_8_CB4	Rubisco small subunit
7009	15.6	6.39	3293	Nano LC-MS-MS	16137614	PPP_sd_8_CB4	Rubisco small subunit
4203	25.6	5.45	223	Nano LC-MS-MS	18331619	PPP_sd_8_CB4	RuBisCo small subunit
3203	26.6	5.24	462	Nano LC-MS-MS	18335130	PPP_sd_8_CB4	Rubisco small subunit
5102	20.2	5.62	2720	Nano LC-MS-MS	18335130	PPP_sd_8_CB4	Rubisco small subunit
3203	26.6	5.24	462	MALDI-TOF-MS	18335346	PPP_sd_8_CB4	Rubisco small subunit
7307	33.7	6.31	432	Nano LC-MS-MS	18335596	PPP_sd_8_CB4	Rubisco small subunit
7214	24.1	6.37	209	Nano LC-MS-MS	18336004	PPP_sd_8_CB4	RuBisCo small subunit
7404	40.8	6.21	539	MALDI-TOF-MS	18336004	PPP_sd_8_CB4	Rubisco small subunit
7404	40.8	6.21	539	Nano LC-MS-MS	18336004	PPP_sd_8_CB4	Rubisco small subunit
8005	16.6	6.63	519	MALDI-TOF-MS	18336004	PPP_sd_8_CB4	Rubisco small subunit
8307	33.2	6.65	858	MALDI-TOF-MS	18336004	PPP_sd_8_CB4	Rubisco small subunit
6211	25.6	6.07	445	Nano LC-MS-MS	18336220	PPP_sd_8_CB4	Rubisco small subunit
8006	15.1	6.68	5793	Nano LC-MS-MS	18336220	PPP_sd_8_CB4	Rubisco small subunit
8107	20.4	6.73	223	Nano LC-MS-MS	18336220	PPP_sd_8_CB4	RuBisCo small subunit
8203	24.1	6.66	1573	Nano LC-MS-MS	18336220	PPP_sd_8_CB4	Rubisco small subunit
8404	38.8	6.64	4861	Nano LC-MS-MS	18336220	PPP_sd_8_CB4	Rubisco small subunit
3010	15.4	5.22	2504	Nano LC-MS-MS	18336801	PPP_sd_8_CB4	Rubisco small subunit
9004	14.8	6.97	6878	MALDI-TOF-MS	18337269	PPP_sd_8_CB4	Rubisco small subunit
8003	15.3	6.50	3251	Nano LC-MS-MS	18337964	PPP_sd_8_CB4	Rubisco small subunit
8608	55.7	6.64	2507	Nano LC-MS-MS	18337964	PPP_sd_8_CB4	Rubisco small subunit
8005	16.6	6.63	519	Nano LC-MS-MS	18339573	PPP_sd_8_CB4	Rubisco small subunit
7112	21.4	6.35	2205	Nano LC-MS-MS	18339779	PPP_sd_8_CB4	Rubisco small subunit
7213	24.9	6.36	667	Nano LC-MS-MS	18339779	PPP_sd_8_CB4	Rubisco small subunit
8406	36.6	6.65	224	Nano LC-MS-MS	18341427	PPP_sd_8_CB4	RuBisCo small subunit
5703	84.2	5.59	229	Nano LC-MS-MS	18371919	PPP_sd_8_CB4	RuBisCo small subunit
7413	38.6	6.21	1373	Nano LC-MS-MS	18373519	PPP_sd_8_CB4	Rubisco small subunit
2403	40.2	4.95	525	Nano LC-MS-MS	14597721	PPP_sd_104_C1	Sedoheptulose-1,7-bisphosphatase, chloroplast precursor
2409	40.2	5.14	311	Nano LC-MS-MS	7584111	PPP_4571_C1	Sedoheptulose-1,7-bisphosphatase, chloroplast precursor
3007	15.0	5.31	1335	Nano LC-MS-MS	18333772	PPP_6838_C1	Thioredoxin m
3010	15.4	5.22	2504	Nano LC-MS-MS	15728594	PPP_6838_C1	Thioredoxin m
3310	27.6	5.30	350	Nano LC-MS-MS	14597777	PPP_sd_80_C1	Triosephosphate isomerase, chloroplast precursor
Functional category 3: cell growth/division							
1614	61.7	4.46	236	Nano LC-MS-MS	18349309	PPP_8378_C1	Nucleosome/chromatin assembly factor group A
Functional category 4: transcription							
6107	20.2	6.03	621	Nano LC-MS-MS	[P10979]	nd	Glycine-rich RNA-binding, abscisic acid inducible
6108	19.3	6.03	611	Nano LC-MS-MS	[P10979]	nd	Glycine-rich RNA-binding, abscisic acid inducible
6107	20.2	6.03	621	Nano LC-MS-MS	18348922	PPP_sd_54_C1	Glycine-rich RNA-binding protein
6108	19.3	6.03	611	Nano LC-MS-MS	18360900	PPP_sd_54_C1	Glycine-rich RNA binding protein
7406	37.4	6.27	862	Nano LC-MS-MS	7583577	PPP_2257_C1	mRNA binding protein precursor
3303	28.3	5.18	712	Nano LC-MS-MS	18328956	PPP_11264_C1	Protein At2g37520, chloroplast precursor
8409	41.7	6.84	274	MALDI-TOF-MS	7287002	PPP_588_C1	Putative RNA-binding protein
8409	41.7	6.84	274	Nano LC-MS-MS	18335549	PPP_588_C1	g5bf protein
8417	38.8	6.77	222	Nano LC-MS-MS	18335549	PPP_588_C1	Putative RNA-binding protein
8419	37.5	6.83	664	Nano LC-MS-MS	18335549	PPP_588_C1	Putative RNA-binding protein
8419	37.5	6.83	664	Nano LC-MS-MS	18360493	PPP_588_C1	Putative RNA-binding protein
Functional category 5: protein synthesis							
2513	44.4	5.10	433	Nano LC-MS-MS	18327036	PPP_11143_C1	30S ribosomal protein S1, chloroplast precursor
2513	44.4	5.10	433	MALDI-TOF-MS	18327036	PPP_11143_C1	30S ribosomal protein S1, chloroplast precursor

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
3504	44.3	5.18	155	Nano LC-MS-MS	18327036	PPP_11143_C1	30S ribosomal protein S1, chloroplast precursor (CS1)
3504	44.3	5.18	155	Nano LC-MS-MS	18341631	PPP_4228_C1	30S ribosomal protein S1
2208	22.6	5.14	194	Nano LC-MS-MS	7047243	PPP_863_C4	50 S chloroplast ribosomal protein L12
9106	19.1	7.00	6627	Nano LC-MS-MS	18325496	PR-AW599782	60S ribosomal protein L12
6707	70.6	5.98	615	MALDI-TOF-MS	15261763	PPP_87_C1	60S ribosomal protein L17, Strong similarity to
6408	36.0	6.07	168	MALDI-TOF-MS	Q98G93	nd	Peptide chain release factor 1
9106	19.1	7.00	6627	Nano LC-MS-MS	18364297	BG361571	Peptidyl-prolyl cis-trans isomerase
1012	14.2	4.50	4478	Nano LC-MS-MS	7067164	AW126617	Ribosomal protein S21
2005	13.8	5.06	1136	Nano LC-MS-MS	7067164	AW126617	Ribosomal protein S21
Functional category 6: protein destination and storage							
2502	50.9	4.83	226	Nano LC-MS-MS	15320220	PPP_11573_C1	"F20B17.8"; [imported]
5003	13.8	5.67	300	Nano LC-MS-MS	18353714	PPP_4829_C1	10 kDa chaperonin
4407	37.3	5.50	456	Nano LC-MS-MS	15327152	PPP_222_C1	60S acidic ribosomal protein
4407	37.3	5.50	456	Nano LC-MS-MS	18334253	PPP_222_C1	60S acidic ribosomal protein
5222	26.3	5.67	316	Nano LC-MS-MS	7207592	PR-AW561639	Chaperonin 21 precursor
5223	25.6	5.68	407	Nano LC-MS-MS	7207592	PR-AW561639	Chaperonin 21 precursor
4611	60.9	5.52	293	MALDI-TOF-MS	[P29197]	nd	Chaperonin CPN60, mitochondrial precursor
5212	26.5	5.79	1110	Nano LC-MS-MS	7207592	PR-AW561639	cp10-like protein (Chaperonin)
8701	80.2	6.42	326	Nano LC-MS-MS	18336740	PPP_1805_C1	Cucumisin-like serine protease, putative
2303	27.4	4.83	330	Nano LC-MS-MS	18327132	PPP_3770_C1	Cysteine proteinase inhibitor
2303	27.4	4.83	330	MALDI-TOF-MS	18345287	PPP_3770_C1	Cysteine proteinase inhibitor
2704	80.0	5.00	1315	Nano LC-MS-MS	18336269	PPP_4091_C1/ PPP_270_C1	hsp 70 protein, putative
5702	70.8	5.59	376	Nano LC-MS-MS	18339274	PPP_975_C1	Heat shock 70 protein
2601	60.3	4.64	255	Nano LC-MS-MS	18343741	PPP_8198_C1	Latex-abundant protein, putative
4605	54.8	5.46	171	Nano LC-MS-MS	18337305	PPP_3522_C1	Mitochondrial processing peptidase, putative
3609	53.6	5.33	681	Nano LC-MS-MS	18339604	PPP_565_C1	Mitochondrial processing peptidase, putative
3607	55.2	5.28	1543	Nano LC-MS-MS	18335009	PPP_565_C1	Mitochondrial processing peptidase, putative
1508	48.4	4.31	382	MALDI-TOF-MS	8272455	PPP_sd_182_C1	Multiubiquitin chain-binding protein
1613	52.7	4.39	703	Nano LC-MS-MS	18372410	PPP_sd_182_C1	Multiubiquitin chain-binding protein
1508	48.4	4.31	382	Nano LC-MS-MS	18375025	PPP_sd_182_C1	Multiubiquitin chain-binding protein
1310	34.7	4.32	1431	Nano LC-MS-MS	7067581	PPP_1383_C2	Nascent polypeptide associated complex α chain
1213	23.8	4.00	3460	Nano LC-MS-MS	15261154	PPP_429_C1	Oxalate oxidase-like germin
8105	21.2	6.55	1358	Nano LC-MS-MS	18364721	PPP_429_C1	Oxalate oxidase-like germin
7407	39.0	6.32	1368	Nano LC-MS-MS	18354139	PPP_7672_C1	Probable protein disulfide-isomerase precursor
3515	51.5	5.23	220	Nano LC-MS-MS	15590637	PPP_4373_C1	Protease regulatory subunit 6A homolog (TAT-binding protein homolog 1)
3207	26.7	5.32	172	Nano LC-MS-MS	[O23708]	nd	Proteasome subunit α type 2
3310	27.6	5.30	350	Nano LC-MS-MS	18347210	PPP_3572_C2	Proteasome subunit α type 2
5221	27.5	5.67	365	Nano LC-MS-MS	18335635	PPP_1642_C2	Proteasome subunit α type 3
2302	27.6	4.71	349	Nano LC-MS-MS	16943776	PPP_sd_75_C1	proteasome subunit α type 5
6211	25.6	6.07	445	Nano LC-MS-MS	18341261	PPP_500_C1	Proteasome subunit β type 1
2708	80.7	5.05	183	Nano LC-MS-MS	[Q02028]	nd	Stromal 70 kDa heat shock-related protein
2410	40.4	4.76	419	Nano LC-MS-MS	18340364	PPP_2709_C1	Thylakoid lumen rotamase, putative
Functional category 7: transporters							
6205	24.7	5.91	630	Nano LC-MS-MS	18329503	BJ161509	Probable H^+-transporting ATPase
Functional category 8: intracellular traffic							
9004	14.8	6.97	6878	Nano LC-MS-MS	7147761	PPP_11472_C1	Core protein – garden pea
5001	14.9	5.61	6871	Nano LC-MS-MS	18344388	PPP_2019_C1	Nuclear transport factor 2
4004	18.1	5.45	1998	Nano LC-MS-MS	18371088	PPP_2527_C1	Pore protein homolog

(*continued on next page*)

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
Functional category 9: cell structure							
3612	55.8	5.35	548	Nano LC-MS-MS	18334571	PPP_3192_C1/ PPP_sd_41_C1	α-Tubulin
3607	55.2	5.28	1543	Nano LC-MS-MS	18324723	PPP_sd_41_C1	α-Tubulin
3610	57.2	5.22	309	MALDI-TOF-MS	[P37832]	nd	β-Tubulin
2615	57.6	5.12	289	Nano LC-MS-MS	[P46263]	nd	β-Tubulin
3610	57.2	5.22	309	Nano LC-MS-MS	[Q41784]	nd	β-Tubulin
2615	57.6	5.12	289	MALDI-TOF-MS	[Q9ZPP0]	nd	β-Tubulin
3610	57.2	5.22	309	Nano LC-MS-MS	18324970	PPP_sd_3_C4	β-Tubulin
3610	57.2	5.22	309	Nano LC-MS-MS	18336312	PPP_sd_3_C4	β-Tubulin
3602	57.6	5.18	557	Nano LC-MS-MS	18337880	PPP_sd_3_C4	β-Tubulin
3602	57.6	5.18	557	Nano LC-MS-MS	18355726	PPP_sd_3_C4	β-Tubulin
2615	57.6	5.12	289	Nano LC-MS-MS	18341760	PPP_sd_3_C4/ PPP_2082_C1	β-Tubulin
2202	26.6	4.63	1837	Nano LC-MS-MS	7207614	PPP_3883_C1	P0019D06.28 protein
5406	37.0	5.65	658	Nano LC-MS-MS	7286315	PPP_74_C1	Pectinesterase, putative
5407	38.3	5.65	386	Nano LC-MS-MS	7286315	PPP_74_C1	Pectinesterase, putative
6402	37.1	5.86	238	Nano LC-MS-MS	7286315	PPP_74_C1	Pectinesterase, putative
5406	37.0	5.65	658	Nano LC-MS-MS	18340504	PPP_74_C1	Pectinesterase, putative
Functional category 10: signal transduction							
8105	21.2	6.55	1358	Nano LC-MS-MS	18334803	PPP_sd_176_C1	Non-symbiotic hemoglobin 0
Functional category 11: disease/defense							
1012	14.2	4.50	4478	Nano LC-MS-MS	15261870	PPP_3965_C1	Copper homeostasis factor, putative
6101	21.3	5.85	886	Nano LC-MS-MS	18346695	PPP_1046_C1	Superoxide dismutase (SOD)
4206	24.6	5.56	2144	Nano LC-MS-MS	7648275	PPP_396_C2	CuZn-superoxide dismutase
4208	24.5	5.42	241	Nano LC-MS-MS	7648275	PPP_396_C2	CuZn-superoxide dismutase
5210	24.8	5.74	1740	Nano LC-MS-MS	7648275	PPP_396_C2	CuZn-superoxide dismutase
7413	38.6	6.21	1373	Nano LC-MS-MS	7648275	PPP_396_C2	CuZn-superoxide dismutase
4201	25.0	5.40	207	Nano LC-MS-MS	15327479	PPP_396_C2	CuZn-superoxide dismutase
4201	25.0	5.40	207	Nano LC-MS-MS	18342624	PPP_396_C2	CuZn-superoxide dismutase
4208	24.5	5.42	241	Nano LC-MS-MS	18342624	PPP_396_C2	CuZn-superoxide dismutase
8104	19.0	6.55	555	Nano LC-MS-MS	18363229	PPP_396_C2	CuZn-superoxide dismutase
6212	24.5	6.07	281	MALDI-TOF-MS	18329031	No hits	Glutathione peroxidase
6212	24.5	6.07	281	Nano LC-MS-MS	18339546	No hits	Glutathione peroxidase
7213	24.9	6.36	667	MALDI-TOF-MS	18327144	PPP_210_C2	Glutathione transferase, putative; protein
7213	24.9	6.36	667	MALDI-TOF-MS	18350097	PPP_210_C2	Glutathione transferase, putative; protein
7213	24.9	6.36	667	MALDI-TOF-MS	18358992	PPP_210_C2	Glutathione transferase, putative; protein
7213	24.9	6.36	667	Nano LC-MS-MS	18362487	PPP_210_C2	Glutathione transferase, putative; protein
7214	24.1	6.37	209	Nano LC-MS-MS	18327144	PPP_210_C2	Glutathione transferase, putative; protein
3304	32.0	5.23	453	MALDI-TOF-MS	12620810	PPP_sd_73_C1	Intracellular pathogenesis-related protein-like protein
3304	32.0	5.23	453	Nano LC-MS-MS	12620810	PPP_sd_73_C1	Intracellular pathogenesis-related protein-like protein
3305	33.5	5.23	504	MALDI-TOF-MS	12620810	PPP_sd_73_C1	Intracellular pathogenesis-related protein-like protein
3305	33.5	5.23	504	Nano LC-MS-MS	12620810	PPP_sd_73_C1	Intracellular pathogenesis-related protein-like protein
3309	34.1	5.29	617	Nano LC-MS-MS	12620810	PPP_sd_73_C1	Intracellular pathogenesis-related protein-like protein
4401	37.5	5.37	495	Nano LC-MS-MS	18353273	PPP_4964_C2	Intracellular pathogenesis-related protein-like protein
5210	24.8	5.74	1740	Nano LC-MS-MS	18364151	No hits	Peroxiredoxin
9007	17.3	7.00	4372	MALDI-TOF-MS	18338291	PPP_6704_C1	Peroxiredoxin, putative
9007	17.3	7.00	4372	Nano LC-MS-MS	18338291	PPP_6704_C1	Peroxiredoxin, putative
Functional category 12: unclear classification							
1413	40.5	4.39	1158	Nano LC-MS-MS	18334175	PPP_3289_C2	Ankyrin-repeat protein HBP1

Table 1 (*continued*)

SSP	$M_r \times 1000$	p*I*	Quantity [counts/IU^2]	Identification via	*P. patens*		Annotation (BLAST search)
					GenBank gi	Cosmoss sequence	
3207	26.7	5.32	172	Nano LC-MS-MS	18337210	PPP_601_C1	NonF-related protein
Functional category 13: unclassified							
2502	50.9	4.83	226	Nano LC-MS-MS	15320220	PPP_11573_C1	"F20B17.8"; [imported]
3108	20.2	5.24	377	Nano LC-MS-MS	7067135	PPP_11578_C1	Hypothetical protein (O. sativa)
3306	29.8	5.25	807	Nano LC-MS-MS	15729123	PPP_2584_C1	Hypothetical 26.4 kDa protein (*O. sativa*)
3511	51.4	5.28	600	Nano LC-MS-MS	18332621	PPP_914_C1	Putative protein; protein id: At4g38220.1 (*A. thaliana*)
3604	53.9	5.22	172	Nano LC-MS-MS	18336034	PPP_601_C1	Conserved hypothetical protein (*Sinorhizobium meliloti*)
4301	30.0	5.38	381	Nano LC-MS-MS	18374725	PPP_7397_C1	SHOOT1 protein [Glycine max]
4301	30.0	5.38	381	MALDI-TOF-MS	18339216	PPP_948_C1	SHOOT1 protein [Glycine max]
4301	30.0	5.38	381	Nano LC-MS-MS	18339216	PPP_948_C1	SHOOT1 protein [Glycine max]
6203	22.6	5.88	189	Nano LC-MS-MS	18333314	PPP_3233_C1	Hypothetical 26.3 kDa (*A. thaliana*)
6205	24.7	5.91	630	Nano LC-MS-MS	18332708	PPP_6923_C1	Unknown protein [*A. thaliana*]
6303	31.0	5.92	222	MALDI-TOF-MS	18367173	PPP_2091_C1	Unknown [*A. thaliana*]gi – 21537242 – gb – AAM61583.1 – [21537242]
6303	31.0	5.92	222	Nano LC-MS-MS	18373772	PPP_2091_C1	Unknown protein [*A. thaliana*]
6303	31.0	5.92	222	Nano LC-MS-MS	18340492	PPP_2091_C1	Unknown protein [*A. thaliana*]
6707	70.6	5.98	615	Nano LC-MS-MS	18368007	PPP_8705_C1	Putative protein (*A. thaliana*)
7102	21.7	6.13	1291	Nano LC-MS-MS	15729178	PPP_1178_C4	Unknown protein [*A. thaliana*]
7505	49.0	6.23	252	Nano LC-MS-MS	18353964	PPP_9061_C1	Hypothetical protein (*A. thaliana*)
8107	20.4	6.73	223	Nano LC-MS-MS	18341740	PPP_1178_C4	Unknown [*A. thaliana*] gi – 21553531 – gb – AAM62624.1 – [21553531]
1209	26.7	4.50	163	Nano LC-MS-MS	18360773	PPP_9487_C1	Not annotated
1212	23.0	4.00	592	Nano LC-MS-MS	18366130	PPP_429_C1	Not annotated
1308	29.6	4.49	2075	Nano LC-MS-MS	18366685	PPP_2462_C1	Not annotated
1308	29.6	4.49	2075	Nano LC-MS-MS	7207406	PPP_2527_C1	Not annotated
1308	29.6	4.49	2075	Nano LC-MS-MS	18374304	PPP_3174_C1	Not annotated
1613	52.7	4.39	703	MALDI-TOF-MS	18352463	PPP_5073_C1	Not annotated
1708	69.0	4.40	404	MALDI-TOF-MS	18344410	PPP_9351_C1	Not annotated
1807	100.9	4.44	354	Nano LC-MS-MS	18339531	PPP_3769_C1	Not annotated
1808	90.6	4.51	655	Nano LC-MS-MS	7584286	AW598980	Not annotated
1808	90.6	4.51	655	Nano LC-MS-MS	18332615	PPP_148_C1	Not annotated
2301	28.9	4.64	288	MALDI-TOF-MS	18352397	PPP_9413	Not annotated
2302	27.6	4.71	349	Nano LC-MS-MS	18355934	PPP_2282_C1	Not annotated
2310	28.8	5.05	758	Nano LC-MS-MS	6167027	AW145346	Not annotated
2402	41.0	4.85	390	MALDI-TOF-MS	18371597	PPP_7838_C1	Not annotated
2613	55.3	5.05	251	MALDI-TOF-MS	18365149	PPP_1740_C1	Not annotated
2615	57.6	5.12	289	MALDI-TOF-MS	18364180	PPP_8540_C1	Not annotated
2804	100.2	4.75	1081	Nano LC-MS-MS	7207709	PPP_273_C1	Not annotated
3102	18.7	5.18	288	Nano LC-MS-MS	18367395	PPP_2527_C1	Not annotated
3102	18.7	5.18	288	MALDI-TOF-MS	18374207	PPP_367_C1	Not annotated
3205	24.6	5.28	1053	Nano LC-MS-MS	18339854	PPP_3035_C1	Not annotated
3209	24.4	5.23	351	Nano LC-MS-MS	18337347	PPP_2150_C1	Not annotated
3310	27.6	5.30	350	MALDI-TOF-MS	18347929	BJ196837	Not annotated
3512	44.6	5.31	254	MALDI-TOF-MS	7207345	PPP_11706_C1	Not annotated
4208	24.5	5.42	241	MALDI-TOF-MS	18349249	PPP_9378_C1	Not annotated
4301	30.0	5.38	381	Nano LC-MS-MS	18326964	PPP_948_C1	Not annotated
4309	29.2	5.49	308	MALDI-TOF-MS	18340515	PPP_6587_C1	Not annotated
4605	54.8	5.46	171	Nano LC-MS-MS	18329434	PPP_11080_C1	Not annotated
6411	40.2	5.93	736	MALDI-TOF-MS	18324881	PPP_11384_C1	Not annotated
7003	13.5	6.19	222	Nano LC-MS-MS	18371480	PPP_8008_C1	Not annotated
7115	22.1	6.25	265	MALDI-TOF-MS	8348627	PPP_2416_C1	Not annotated
7505	49.0	6.23	252	Nano LC-MS-MS	18340130	PPP_747_C1	Not annotated
8011	12.9	6.85	374	Nano LC-MS-MS	18334774	PPP_1408_C1	Not annotated
9006	13.8	6.99	3555	Nano LC-MS-MS	7067554	PPP_11470_C1	Not annotated

Proteins are grouped according to their functions. When a mass spectrum could not be matched to *P. patens* nucleic acid information, a GI number cannot be provided and the corresponding protein from another species is identified by its Swiss-Prot accession number [given in brackets]. In these cases a Cosmoss search was not performed (nd).

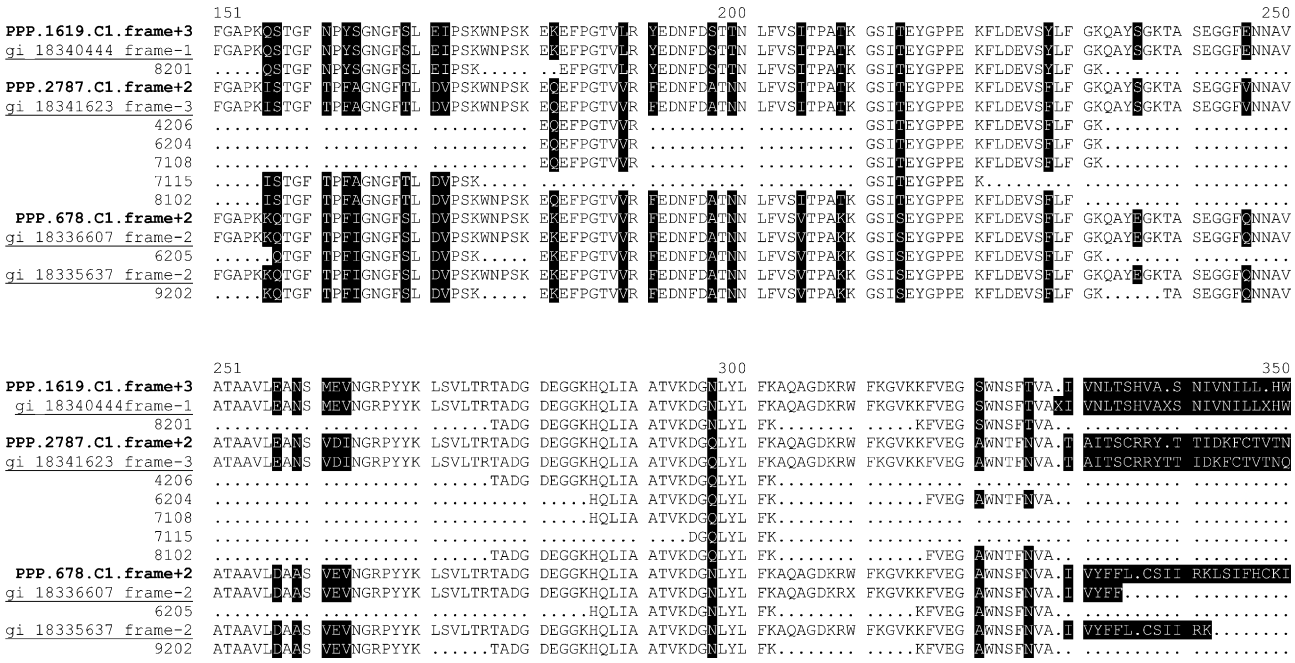

Fig. 5. Alignment of sequenced peptide fragments of eight proteins identified as '23 kDa polypeptide of the oxygen evolving complex' with the deduced amino acid sequences of the corresponding ESTs (underlined) and Cosmoss contigs (bold). Proteins are identified by their Standard Spot Numbers. Only the part of sequence covering all identified peptides is displayed. Positions showing sequence variations are highlighted.

murine and *Escherichia coli* 2D maps and provides only one gel of plant origin. While the search for new drug targets is definitely an aim that fuels progress in the field of human and mammalian proteomics (Hanash et al., 2002; Jeffery and Bogyo, 2003), the lack of reports on plant proteomics is also due to its low amenability.

Extraction of proteins from plant tissues is rather challenging when compared to animal or bacterial samples. Among the reasons are a higher mechanic resistance of the material itself due to the presence of a cell wall that frequently undergoes secondary thickening and rigidification, a relatively low protein content and usually a high amount of endogenous proteases and interfering compounds (e.g., phenolic compounds, carbohydrates, organic acids, pigments, terpenes, inhibitory ions) (Granier, 1988; Gegenheimer, 1990). These obstacles frequently render plant proteomics a rather gruelling enterprise. In order to make *P. patens* protonema accessible to proteomics research, several aspects of standard extraction and electrophoresis procedures had to be optimized to suit the requirements for this seed-less plant. Tissue disruption could only successfully be performed employing a ball mill, as the protonema cells turned out to be highly resistant towards mechanical force (Heintz et al., 2004). Inclusion of thiourea in the IEF sample buffer, though recommended to improve protein extraction from other tissues (Rabilloud et al., 1997), had a devastating effect on the performance of isoelectric focusing in the case of Physcomitrella and had to be omitted (results not shown). For the same reason, the sequential protein solubilization procedure for plant proteins developed by Jacobs et al. (2001) could not improve the performance of protein display. The pH of the equilibration buffer was raised to 8.3 (matching the pH of the SDS running buffer rather than the pH of the stacking gel) to optimize conditions for reduction and alkylation (Herbert et al., 2001). In order to improve the performance of 2D electrophoresis, tributylphosphine (TBP) rather than dithiothreitol (DTT) was included as a reducing agent both in the IEF sample buffer and the IPG strip equilibration buffer (Herbert et al., 1998). For the same reason, the crosslinker bisacrylamide was replaced with piperazine diacrylamide (PDA) in the SDS gels (Hochstrasser et al., 1988). Severe vertical streaking of the proteins during SDS electrophoresis could be remedied by increasing the amount of SDS in both the IPG strip equilibration buffer (to 4%) and in the SDS–PAGE running buffer (to 0.4%).

Proteomic relies heavily on the availability of profound protein or gene databases. Currently, plant genome sequencing programs have only been completed for *A. thaliana* and *Oryza sativa*. These are the model organisms that have therefore been in the main focus of plant proteomics. There are, however, currently non-redundant datasets or gene indices for an additional 18 plant species available at the TIGR website (http://www.tigr.org/tdb/tgi/plant.shtml), including

wheat, corn, soybean, barley and rye. The Sputnik resource for the annotation of clustered plant ESTs (http://mips.gsf.de/proj/sputnik/) currently lists 28 ''genomeless genomes'' from plants covering a range from Chlamydomonas to poplar. Apart form the model organisms Arabidopsis and rice, publications in the field of plant proteomics have also addressed plants as wheat (Andon et al., 2002), barley (Østergaard et al., 2002), corn (Porubleva et al., 2001), tomato (Rep et al., 2002), pea (Bardel et al., 2002), tobacco (Blee et al., 2001), Medicago (Mathesius et al., 2001), clover (Wilson et al., 2002) Chlamydomonas (van Lis et al., 2003) and members of the Brassicaceae family (Marquès et al., 2001). Concerning Physcomitrella, two-dimensional separation of total proteins from protoplasts has been reported by Doonan and Duckett (1988). The protein pattern was further probed with antibodies to detect α and β tubulin subunits. The proteome of isolated *P. patens* chloroplasts was investigated more profoundly (Kasten et al., 1997) when following two-dimensional electrophoresis sequences from 10 protein spots were obtained by N-terminal and internal Edman sequencing. More recently, a sensitive method suitable for the detection of protein phosphorylation in Physcomitrella has been described by Heintz et al. (2004).

Recently, *P. patens* has been described as an organism suitable for molecular farming, i.e., the production of pharmaceutically important proteins following genetic engineering (Decker et al., 2003). This emerging field of biotechnology is expected to benefit substantially from further progress in *P. patens* proteomics.

3. Concluding remarks

The seed-less plant *P. patens* is the only land plant known to allow for gene targeting via homologous recombination. The moss is therefore regarded as the ideal model organism in the field of plant functional genomics. We now demonstrate its feasibility towards the emerging field of proteomics by providing a first map of Physcomitrella protonema proteins. These results are expected to add a new quality to the ongoing worldwide research on *P. patens* and will thereby contribute considerably to progress in plant functional genomics.

4. Experimental

4.1. Sources of chemicals

Acrylamide, piperazine diacrylamide (PDA), Biolytes 3-10 and 20% sodium dodecyl sulphate (SDS) were purchased from Bio-Rad (Hercules, CA, USA). Iodoacetamide, tributylphosphine (TBP) and trichloroacetic acid (TCA) were from Fluka (Buchs, Swizerland). Sequencing grade modified trypsin was from Promega (Madison, WI, USA). Dithiothreitol (DTT) was purchased from Roche Diagnostics (Mannheim, Germany). 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS) and Coomassie Brillant Blue G250 were from Roth (Karlsruhe, Germany). α-Cyano-4-hydroxycinnamic acid and urea (electrophoresis grade) were from Sigma–Aldrich (St. Louis, MO, USA).

All other chemicals were of reagent grade. All solutions were prepared with HPLC-grade (bidistilled) water.

4.2. Growth of plant material

P. patens (Hedw.) B.S.G. protonema was grown in 5 litre semi-continuous bioreactors in Knop medium with the pH controlled at 4.5 as described (Hohe et al., 2002) at a density of 900 mg fresh weight per litre. Plant material was collected from the medium via paper filtration in a buchner funnel with suction, immediately frozen in liquid nitrogen and stored at – 80 °C until use.

4.3. Extraction of total proteins

Frozen protonema was ground to a fine powder in a ball mill (Mikro-Dismembrator S, B. Braun Biotech International, Melsungen, Germany) for one minute at 2600 rpm. PFTE flasks and grinding balls were precooled in liquid nitrogen in order to avoid thawing of the sample during the grinding process. Total proteins were extracted with TCA/acetone as described (Damveral et al., 1986). Briefly, 300 mg of frozen material were extracted with 1.5 ml ice cold 10% TCA in acetone supplemented with 13 mM DTT. Samples were allowed to stand at −20 °C for 1 h and were subsequently centrifuged at 19,000*g* for 15 min at −5 °C. The pellet was washed three times with 1.5 ml of ice cold acetone with 13 mM DTT, left at −20 °C for 1 h between washes and centrifuged as described. The final pellet that was deprived of chlorophyll was dried in a speedvac and the total proteins were extracted with 600 µl of IEF sample buffer (9.5 M urea, 2% [w/v] CHAPS, 0.2% [w/v] Biolyte 3–10, 25 mM TBP) with shaking at 37 °C for 30 min. Cell debris was removed by centrifuging twice at 19,000*g* at room temperature for 15 min. Protein content of the supernatant was determined using a modified Bradford protocol (Ramagli and Rodriguez, 1985) Aliquoted samples were stored at −80 °C until use.

4.4. First dimension (IEF)

Protein samples equivalent to 200 µg for analytical gels and 1000 µg for preparative gels respectively were

diluted to a final volume of 350 μl with IEF sample buffer and isoelectric focussing was performed using the Protean® IEF Cell (Bio-Rad, Hercules, USA) strictly following the manufacturers instruction. IPG strips 17 cm in length with a linear pH gradient from 4.0 to 7.0 were rehydrated actively with the sample at 50 V and 20 °C for 14 h. Subsequently, the strips were transferred to clean tray channels and isoelectric focussing was performed with a current limit of 50 μA per strip and a voltage limit of 10,000 V at a constant temperature of 20 °C for 80,000 V h.

4.5. Second dimension (SDS–PAGE)

In the second dimension, the proteins were separated on a SDS gel of 18 cm × 18 cm × 1 mm with a linear gradient from 8.0% to 16.7% acrylamide in the running gel and a 1 cm stacking gel with 4% acrylamide. PDA rather than bisacrylamide was used as a crosslinker in all experiments.

Following isoelectric focussing, IPG strips were equilibrated in 10 ml of equilibration buffer (50 mM Tris–HCl, pH 8.3, 4% [w/v] SDS, 6 M urea, 30% [w/v] glycerol) supplemented with 2% (v/v) TBP for 12 $\frac{1}{2}$ min and subsequently in equilibration buffer without TBP but containing 2.5% (w/v) iodoacetamide for the same time. IPG strips were briefly rinsed with water, placed on top of the SDS gel and covered with 1% agarose in SDS running buffer (containing ~25 mM Tris base, 192 mM glycine, 0.4% (w/v) SDS, 0,02% anhydrous sodium thiosulfate). SDS–PAGE was performed using the Protean® II xi cell (Bio-Rad). The upper buffer chamber was filled with 350 ml SDS running buffer while the lower buffer chamber was filled with 1.5 litre of 25 mM Tris–HCl, pH 8.3. Electrophoresis was performed at a constant current of 11 mA/gel with cooling to 15 °C for 14 h. Following electrophoresis, gels were stained with 0.08% Coomassie Brilliant Blue G250 in 20% (v/v) methanol, 7.8% (w/v) ammonium sulphate, 1.33% (w/v) phosphoric acid and destained with water. Two-dimensional protein patterns were documented using the FluorS multiimager (Bio-Rad, Hercules, USA) and analysed with PDQuest software Version 7.1 (BioRad). Excision of proteins from preparative gels was performed using the ProteomeWorks™ spot cutter (Bio-Rad) under the control of the PDQuest software.

4.6. Preparation of samples for mass spectrometry

In gel digestion with trypsin was performed according to published methods (Jenö et al., 1995; Shevchenko et al., 1996) modified for use with a robotic digestion system (MassPREP, Micromass, Manchester, UK). Gel pieces were decolorized by washing with 50 μL of 25 mM ammonium hydrogen carbonate followed by washing with 50 μL of acetonitrile. This operation was repeated twice. Cysteine residues were reduced with DTT and derivatized by treatment with iodoacetamide. The gel pieces were then dehydrated with acetonitrile and dried at 60 °C, prior to addition of modified trypsin (10 μL at 12.5 ng/μL in 25 mM ammonium hydrogen carbonate). The digestion was performed at 35 °C overnight. 5 μL of 30% water/65% acetronitrile/5% formic acid were added in each gel pieces.

4.7. MALDI mass spectrometry

Mass measurement were carried out on a Biflex II (Bruker, Wissenbourg, F) matrix assisted laser desorption time of flight mass spectrometer (MALDI-TOF). A saturated solution of α-cyano-4-hydroxycinnamic acid in acetone was used as a matrix. A first layer of fine matrix crystals was obtained by spreading and fast evaporation of 0.5 μL of matrix solution. On this fine layer of crystals, a droplet of 0.5 μL of aqueous formic acid (5%) solution was deposited. Afterwards, 0.5 μL of sample was added and a second 0.25 μL droplet of saturated matrix solution (in 50% water / 50% acetonitrile) was added. The preparation was dried under vacuum. The sample was washed by applying 1 μL of formic acid (5%) solution.

All mass spectra were internally calibrated with trypsin autolysis peaks. The resulting peptide mass fingerprints (PMFs) were searched against a local copy of the non redundant database SWISS-PROT (http://www.expasy.ch/sprot) using MASCOT (Perkins et al., 1999) search program. The parameters used in the search were as follows: peptide mass tolerance 50 ppm, 1 missed cleavage, carboxymethylated cysteine, methionine oxidation and N-terminal acetylation.

4.8. Liquid chromatography tandem mass spectrometry

Sample were injected into a CapLC (Waters, Milford, MA, USA) System equipped with an autosampler, gradient and auxiliary pump. 6.4 μL was injected via "microliter pickup" mode and desalted on-line through a 300 μm × 5 mm C_{18} trapping cartridge (LC Packings, San Fransisco, CA, USA). The samples were desalted at high flow (30 μl/min) for 3 min. The peptides were separated on a 75 μm × 15 cm × 3 μm C_{18} 100 Å PepMap™ column (LC Packings, CA, USA) prior to introduction into the mass spectrometer. A typical reversed-phase was used from low to high organic over about 35 min. Mobile phase A was 0.1% formic acid and B was 95% acetronitrile, 0.1% formic acid. The flow rate was 5 μL/min. The system utilized a split flow resulting in a column flow rate of approximately 400–500 nL/min.

MS/MS data were obtained using a Q-Tof 2 (Micromass, Manchester, UK) fitted with Z-spray nanoflow electrospray ion source. The mass spectrometer was operated in positive ion mode with a potential of 3500V applied to the nanoflow probe body. The collision energy was determined on the fly based on the mass and charge state of the peptides. Charge state recognition was used to switch only doubly and triply charged ions into MS-MS mode. Several trypsin autolysis ions were excluded. The data were processed by Protein Lynx Version (Micromass, Manchester, UK) to generate searchable peak lists. Initial protein identification were made by correlation of uninterpreted tandem mass spectra to entries in SWISS-PROT using Global Server (Version 1.1, Micromass).

In addition, if clearly separated proteins were assigned to different Physcomitrella ESTs that could be traced to identical protein functions (or no function at all) an additional BLAST search of the sequence data against the clustered EST database Cosmoss (at http://www.cosmoss.org/bm/) (Rensing et al. in preparation) was performed in order to ascertain whether or not the different sequences were covered by one and the same contig.

Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft (DFG Re 387/7) is gratefully acknowledged.

References

Andon, N.L., Hollingworth, S., Koller, A., Greenland, A.J., Yates III, J.R., Haynes, P.A., 2002. Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. Proteomics 2, 1156–1168.

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.

Beranova-Giorgianne, S., 2003. Proteome analysis by two-dimensional gel electrophoresis and mass spectrometry: strength and limitations. Trends Anal. Chem. 22, 273–281.

Berg, P., 1991. Reverse genetics: its origins and prospects. BioTechnology 9, 342–344.

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.

Blee, K.A., Wheatley, E.R., Bonham, V.A., Mitchell, G.P., Robertson, D., Slabas, A.R., Burrell, M.M., Wojtaszek, P., Bolwell, G.P., 2001. Proteomic analysis reveals a novel set of cell wall proteins in a transformed tobacco cell culture that synthesises secondary walls as determined by biochemical and morphological parameters. Planta 212, 404–415.

Bouche, N., Bouchez, D., 2001. Arabidopsis gene knockout: phenotypes wanted. Curr. Opin. Plant Biol. 4, 111–117.

Bronson, S.K., Smithies, O., 1994. Altering mice by homologous recombination embryonic stem cells. J. Biol. Chem. 269, 27155–27158.

Chiari, M., Manzochhi, A., Righetti, P.G., 1990. Formation of a cysteine-acrylamide adduct in isoelectric focusing gels. J. Chromatogr. A 500, 697–704.

Damveral, C., de Vienne, D., Zivy, M., Thiellement, H., 1986. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis 7, 52–54.

Decker, E.L., Gorr, Gorr Reski, R., 2003. Moss – an innovative tool for protein production. BioForum Eur. 7, 96–97.

Doonan, J.H., Duckett, J.G., 1988. The bryophyte cytoskeleton: experimental and immunofluorescence studies on morphogenesis. Adv. Bryol. 3, 1–31.

Egener, T., Granado, J., Guitton, M.-C., Hohe, A., Holtorf, H., Lucht, J.M., Rensing, S.A., Schlink, K., Schulte, J., Schween, G., Zimmermann, S., Duwenig, E., Rak, B., Reski, R., 2002. High frequency of phenotypic deviations in *Physcomitrella patens* plants transformed with a gene disruption library. BMS Plant Biol. 2, 6. Available from <http://www.biomedcentral.com/1471-2229/2/6>.

EU Arabidopsis Genome Project, 1998. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391, 485–488.

Fey, S.J., Larsen, P.M., 2001. 2D or not 2D. Curr. Opin. Chem Biol. 5, 26–33.

Fountoulakis, M., Juranville, J.-F., Berndt, P., Lanken, H., Suter, L., 2001. Two-dimensional database of mouse liver proteins. An update. Electrophoresis 22, 1747–1763.

Gegenheimer, P., 1990. Preparation of extracts from plants. In: Deutscher, M.P. (Ed.), Methods in Enzymology, vol. 182. Academic Press, San Diego, pp. 174–193.

Gooley, A.G., Packer, N.H., 1997. The importance of proteins co- and post-translational modifications in proteome projects. In: Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.), Proteome Research: New Frontiers in Functional Genomics. Springer, Berlin, pp. 65–86.

Görg, A., Postel, W., Günther, S., 1988. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 9, 531–546.

Granier, F., 1988. Extraction of plant proteins for two-dimensional electrophoresis. Electrophoresis 9, 712–718.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R., 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc. Natl. Acad. Sci. USA 97, 9390–9395.

Hanash, S.M., Madoz-Gurpide, J., Misek, D.E., 2002. Identification of novel targets for cancer therapy using expression proteomics. Leukemia 16, 478–485.

Hari, V., 1981. A method for the two-dimensional electrophoresis of leaf proteins. Anal. Biochem. 113, 332–335.

Heermeier, K., Prydz, H., Reinhard-Rupp, J., Engels, J., 2002. Meeting review: ESF workshop on 'Impact of nucleic acid chemistry on gene function analysis: antisense, aptamers, ribozymes and RNAi'. Comp. Funct. Genom. 3, 441–446.

Heintz, D., Wurtz, V., Van Dorsselaer, A., Reski, R., Sarnighausen, E., 2004. An efficient protocol for the identification of protein phosphorylation in a seedless plant, sensitive enough to detect members of signalling cascades. Electrophoresis 25, 1149–1159.

Hellwege, E.M., Dietz, K.-J., Hartung, W., 1996. Abscisic acid causes changes in gene expression involved in the induction of the landform of the liverwort *Riccia fluitans* L. Planta 198, 423–432.

Hille, J.M., Freed, A.L., Wätzig, H., 2001. Possibilities to improve automation, speed and precision of proteome analysis: a comparison of two-dimensional electrophoresis and alternatives. Electrophoresis 22, 4035–4052.

Henning, S., Beste, G., 2002. Loss-of-function strategies in drug target validation. Curr. Drug Discov. (May), 17–21. Available from <http://www.currentdrugdiscovery.com>.

Herbert, B.R., Molloy, M.P., Gooley, A.A., Walsh, B.J., Bryson, W.G., Williams, K.L., 1998. Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. Electrophoresis 19, 845–851.

Herbert, B., Galvani, M., Hamdan, M., Olivieri, E., MacCarthy, J., Pedersen, S., Righetti, P.G., 2001. Reduction and alkylation of proteins in preparation of two-dimensional map electrophoresis: why, when, and how? Electrophoresis 22, 2046–2057.

Hochstrasser, D.F., Harrington, M.G., Hochstrasser, A.C., Miller, M.J., Merril, C.R., 1988. Methods for increasing the resolution of two-dimensional protein electrophoresis. Anal. Biochem. 173, 424–435.

Hohe, A., Reski, R., 2002. Optimisation of a bioreactor culture of the moss *Physcomitrella patens* for mass production of protoplasts. Plant Sci. 163, 69–74.

Holtorf, H., Guitton, M.-C., Reski, R., 2002. Plant functional genomics. Naturwissenschaften 89, 235–249.

Hurkman, W.J., Tanaka, C.K., 1986. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol. 81, 802–806.

Jacobs, D.I., van Rijssen, M.S., van der Heijden, R., Verpoorte, R., 2001. Sequential solubilization of proteins precipitated with trichloroacetic acid in acetone from cultured *Catharantus roseus* cells yields 52% more spots after two-dimensional electrophoresis. Proteomics 1, 1345–1350.

Jeffery, D.A., Bogyo, M., 2003. Chemical proteomics and its application to drug discovery. Curr. Opin. Biotechnol. 14, 87–95.

Jenö, P., Mini, T., Moes, S., Hintermann, E., Horst, M., 1995. Internal sequences from proteins digested in polyacrylamide gels. Anal. Biochem. 224, 75–82.

Kasten, B., Bock, F., Nuske, J., Reski, R., 1997. Cytokinin affects nuclear- and plastome-encoded energy converting plastid enzymes. Planta 201, 261–272.

Koller, A., Washburn, M.P., Lange, M.B., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., Yates III, J.R., 2002. From the Cover: proteomic survey of metabolic pathways in rice. Proc. Natl. Acad. Sci. USA 99, 11969–11974.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., Rehrauer, W.M., 1994. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. 58, 401–465.

Langen, H., Berndt, P., 2001. Proteomics databases. In: James, P. (Ed.), Proteome Research: Mass Spectrometry. Springer, Berlin, pp. 229–257.

Levanon, D., Brenner, O., Otto, F., Groner, Y., 2003. *Runx3* knockouts and stomach cancer. The challenge of identifying phenotypic defects directly attributable to loss of gene function. EMBO Rep. 4, 560–564.

van Lis, R., Atteia, A., Mendoza-Hernández, G., González-Halphen, D., 2003. Identification of novel mitochondrial protein components of *Chlamydomonas reinhardtii*. A proteomic approach. Plant Physiol. 132, 318–330.

Marquès, K., Sarazin, B., Chané-Favre, L., Zivy, M., Thiellement, H., 2001. Comparative proteomics to establish genetic relationships in the Brassicaceae family. Proteomics 1, 1457–1462.

Mathesius, U., Keijzers, G., Natera, S.H.A., Weinman, 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.

McCarthy, J., Hopwood, F., Oxley, D., Laver, M., Castagna, A., Righetti, P.G., Williams, K., Herbert, B., 2003. Carbamylation of proteins in 2-D electrophoresis: myth or reality? J. Proteome Res. 2, 239–242.

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.

Nishiyama, T., Hiwatashi, Y., Sakakibara, I., Kato, M., Hasebe, M., 2000. Tagged mutagenesis and gene trap in the moss *Physcomitrella patens* by shuttle mutagenesis. DNA Res. 7, 9–17.

Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., Hasebe, M., 2002. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc. Natl. Acad. Sci. USA 100, 8007–8012.

Østergaard, O., Melchior, S., Roepstorff, P., Svensson, B., 2002. Initial proteome analysis of mature barley seeds and malt. Proteomics 2, 733–739.

Perkins, D., Pappin, D.J., 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 sequence and expressed sequence tag data for identification of proteins with peptide mass fingerprints. Electrophoresis 22, 1724–1738.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G., 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat. Biotechnol. 19, 45–50.

Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., 1997. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18, 307–316.

Ramagli, L.S., Rodriguez, L.V., 1985. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. Electrophoresis 6, 559–563.

Rensing, S.A., Rombauts, S., Van de Peer, Y., Reski, R., 2002. Moss transcriptome and beyond. Trends Plant Sci. 7, 535–538.

Rep, M., Dekker, H.L., Vossen, J.H., de Boer, A.D., Houterman, P.M., Speijer, D., Back, J.W., de Koster, C.G., Cornelissen, B.J.C., 2002. Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. Plant Physiol. 130, 904–917.

Reski, R., 1998. Development, genetics and molecular biology of mosses. Bot. Acta 111, 1–15.

Reski, R., 2003. *Physcomitrella patens* as a novel tool for plant functional genomics. In: Vasil, I.K. (Ed.), Plant Biotechnology 2002 and Beyond. Kluwer Academic Publishers, Dordrecht, pp. 205–209.

Schaefer, D., Zrÿd, J.-P., 1997. Efficient gene targeting in the moss *Physcomitrella patens*. Plant J. 11, 1195–1206.

Shevchenko, A., Wilm, M., Vorm, O., Mann, M., 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68, 850–858.

Sperling, K., 2001. From proteomics to genomics. Electrophoresis 22, 2835–2837.

Struhl, K., 1983. The new yeast genetics. Nature 305, 391–397.

Tissier, A., Bourgeois, P., 2001. Reverse genetics in plants. Curr. Genom. 2, 269–284.

Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S.J., Larsen, P.M., Görg, A., 2000. Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. Electrophoresis 21, 2610–2616.

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.