Journal of Biochemistry and Molecular Biology, Vol. 38, No. 6, November 2005, pp. 650-660

JOURNAL OF Biochemistry and Molecular Biology

Proteomic Analysis and Extensive Protein Identification from Dry, Germinating Arabidopsis Seeds and Young Seedlings

Qiang Fu[‡], Bai-Chen Wang[‡], Xiang Jin[‡], Hong-Bing Li[‡], Pei Han[‡], Kai-hua Wei[†], Xue-min Zhang[†] and Yu-Xian Zhu[‡,#,*]

[‡]The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, China

[†]Institute of Basic Medical Sciences, National Center of Biomedical Analysis, 27 Tai-Ping Road, Beijing 100850, China

[#]National Plant Gene Research Center (Beijing), Beijing, China 100101

Received 15 April 2005, Accepted 17 August 2005

Proteins accumulated in dry, stratified Arabidopsis seeds or young seedlings, totaled 1100 to 1300 depending on the time of sampling, were analyzed by using immobilized pH gradient 2-DE gel electrophoresis. The molecular identities of 437 polypeptides, encoded by 355 independent genes, were determined by MALDI-TOF or TOF-TOF mass spectrometry. In the sum, 293 were present at all stages and 95 were accumulated during the time of radicle protrusion while another 18 appeared in later stages. Further analysis showed that 226 of the identified polypeptides could be located in different metabolic pathways. Proteins involved in carbohydrate, energy and amino acid metabolism constituted to about 1/4, and those involved in metabolism of vitamins and cofactors constituted for about 3% of the total signal intensity in gels prepared from 72 h seedlings. Enzymes related to genetic information processing increased very quickly during early imbibition and reached highest level around 30 h of germination.

Keywords: 2-DE, Gene ontology, Mass spectrometry, Pathways, Protein identification

Introduction

Germination is the first and may be the foremost growth stage in the plant's life cycle since embryos formed by sexual reproduction assumed the shape and functional characteristics of higher plants during this process. For many years, scientists strive to elucidate the molecular and cellular mechanisms that changed seeds, a resting organ with only negligible levels of metabolism, into actively growing plants (Koornneef *et al*., 1994; Bewley, 1997; Koornneef *et al*., 2002). Disclosure of the Arabidopsis genome sequence (Arabidopsis Genome Initiative, 2000) increased markedly our awareness of the great complexity in regulating plant growth and development by providing the molecular basis for microarray analysis (Bove *et al*., 2001). Many highly expressed, yet functionally unknown genes were identified by screening the arrays with cDNA samples prepared from germinating Arabidopsis seeds of different stages (Girke *et al*., 2000). For example, Yamauchi and colleagues reported that a subset of gibberellin (GA) biosynthesis genes were significantly upregulated in response to low temperature treatment, a condition designed to break seed dormancy and to improve frequency of germination, resulting in an increase in the levels of bioactive GAs and transcript abundance of GA-inducible genes (Yamauchi *et al*., 2004). Rajjou *et al*. (2004) reported that inhibition of protein synthesis with cycloheximide blocked germination completely whereas inhibition of transcription only delayed it. Their data suggested that mRNAs formed and stored in the maturation-drying phase fulfilled all requirements for germination initiation. They concluded that *de novo* protein syntheses, especially syntheses of enzymes involved in reserve mobilization and resumption of various metabolic activities, were necessary for germination to occur.

Among the approximately 1300 Arabidopsis seed proteins detected on 2D gel electrophoresis, 74 were found to change abundance significantly during early imbibition or radicle protrusion (Gallardo *et al*., 2001). However, only 12 of these 74 polypeptides were found to possess some regulatory functions with one each of the WD-40 repeat protein, EM-like protein and luminal-binding protein (LBP), two jasmonate-inducible proteins (JIP), three initiation/elongation factors (IF/

*To whom correspondence should be addressed.
Tel: 86-10-6275-1848; Fax: 86-10-6275-4427
E-mail:zhuyx@water.pku.edu.cn

EF) for genetic information processing (GIP) and four heat-shock proteins (HSP). The rest were either structural- and storage proteins or enzymes involved in reserve mobilization. In a similar study, the same group of scientists found that the cytoskeleton component α-tubulin, one cell wall loosening enzyme β-glucosidase and two isoforms of S-adenosyl-methionine synthetase were highly related to germination (Gallardo *et al.*, 2002). In fact, most currently identified genes and proteins, except *AtGA3ox1* and the DELLA subfamily in the GRAS family of plant transcription regulators involved in GA signal transduction (Yamauchi *et al.*, 2004; Tyler *et al.*, 2004), are down-stream components, rather than upstream regulators of this important process. Also, the number of characterized proteins related to germination is not great enough to allow construct of various metabolic pathways according to their time of activation and to execute detailed kinetic analyses. This kind of information may be vital to our understanding of the molecular mechanisms controlling the very first steps in the life cycle of higher plants. Here we prepared 2-demensional protein gels from non-dormant dry seeds to three days after stratification and continued to 30, 48, 72 and 96 h after germination initiation. We identified a large number of proteins and located 226 of them in various metabolic pathways that were activated at early stages of seed germination.

Material and Methods

Plant growth and sample preparation Dry mature Arabidopsis seeds (Columbia ecotype) were soaked in dark in ddH_2O at 4°C for 3 days before being transferred onto solid MS media and grown under constant temperature (20-22°C) with a light regime of 16/8 h (light/dark cycle). Seeds that showed no elongated radicles 30 h after beginning of the germination experiments were thrown out and were excluded from any further analyses. Protein samples of various growth stages were prepared from dry seeds, 0 h (harvested upon completion of the dark stratification) and 30, 48, 72 or 96 h after beginning of the germination experiments.

Protein extraction All samples (250 mg equivalent of dry or 0 h seeds, or young seedlings harvested at 30 and 48 h, or 1 g of 72 and 96 h seedlings) were ground in liquid nitrogen in a mortar and pestle. The fine powder was precipitated at −20°C with10% (w/v) trichloroacetic acid (TCA) in cold acetone containing 0.07% (w/v) 2-mercaptoethanol for at least 2 h. The mixture was centrifuged at 40,000 g at 4°C for 1 h, and the precipitates were washed with cold acetone containing 0.07% (w/v), 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 2 mM EDTA. Pellets were dried by vacuum centrifugation and solubilized in 7 M urea, 4% (w/v) CHAPS, 20 mM DTT, 0.1% pharmalyte 3-10 (Amersham Pharmacia Biotech, Uppsala, Sweden), 1% (v/v) proteinase-inhibitor mixture (Amersham Pharmacia Biotech) at room temperature for 1 h and then the mixture was centrifuged at 4°C for 1 h. The supernatant was stored at −80°C until further use. In general, about 10-20 mg total proteins were obtained from each sample when quantified using the 2-D Quant kit (Amersham Pharmacia Biotech).

Two-dimensional electrophoresis and quantitative analyses 2-DE was performed as described by Gorg *et al.* (1988). Isoelectric focusing was done with the IPGphor system (Amersham Pharmacia Biotech). Immobiline 3-10 and 4-7 linear DryStrips (Amersham Pharmacia Biotech) were run at 30 V for 8 h, 50 V 4 h, 100 V 1 h, 300 V 1 h, 500 V 1 h, 1000 V 1 h and 8000 V 12 h using rehydration buffer (8 M urea, 2% CHAPS, 20 mM dithiothreitol) and 0.5% IPG buffer (Amersham Pharmacia Biotech). SDS-PAGE was performed using 13% polyacrylamide gels without a stacking gel in the ETTAN DALT SIX ELECT UNIT 230 (Amersham Pharmacia Biotech). Following SDS-PAGE, gels were stained with 0.1% (w/v) Coomassie Blue R-350 (CBR350) in 10% acetic acid and destained with 10% acetic acid or silver staining according to Shevchenko *et al.* (1996). Gel images were processed using 2-D Imagemaster Elit (version 4.01, Amersham Pharmacia Biotech) according to the manufacturer's instruction. The 2-D gels were aligned, matched, and volumes of each spot were quantitatively determined after normalization of total spot volumes and background subtraction.

In-gel digestion In-gel digestion was performed following a previously published protocol (Gamble *et al.*, 2000). Briefly, protein spots were excised from 2-D gels stained with CBR350 and destained by washing in 200 ml aliquots of 50 mM ammonium bicarbonate in 50% v/v acetonitrile for 30 min. Each spot was then dried in a SpeedVac (Savant Instruments, Holbrook, NY) and rehydrated at 4°C for 30 min in 10 ml digestion solution (25 mM ammonium bicarbonate and 0.01 mg/ml sequence-grade modified trypsin) (Roche, Mannheim, Germany), 4-10 ml digestion solution without trypsin was then added to keep it wet during the digestion. The digestion was stopped after overnight incubation at 37°C by immersing the protein spots in 5% trifluoroacetic acid (TFA, MERCK, Honenbrunn, Germany) for 20 min. Peptides were extracted by 20 ml 5% TFA for 1 h at 37°C and then by 20 ml 2.5% TFA/50% acetonitrile for 1 h at 37°C. The combined supernatants were evaporated in the SpeedVac again and dissolved in 4 ml 0.5% aqueous TFA for mass spectrometric analysis.

Peptide mass fingerprinting by MALDI-TOF-MS and TOF-TOF-MS All mass spectra of MALDI-TOF-MS were obtained on an Ultraflex™ MALDI-TOF/TOF (Bruker-Franzen, Bremen, Germany) in positive ion mode at an accelerating voltage of 20 kV with a-cyano-4-hydroxy cinnamic acid as the matrix. The spectra were internally calibrated using trypsin autolysis products. PMFs obtained were used to search the NCBInr database by the Mascot search engine (http://www.matrixscience.co.uk) with a tolerance of ±0.1Da and one missed cleavage site. TOF/TOF MS of selected polypeptides was performed when PMFs obtained from a particular spot did not provide sufficient information for protein identification. TOF/TOF MS analyses were repeated using preparations from the same protein spot for accuracy.

Protein annotation, pathway identification and gene categorization All proteins identified on 2-D gels were first searched against TAIR (The Arabidopsis Information Resource) database (Dec 14, 2004) (Rhee *et al.*, 2003) for functional annotations and also for presence of full-length or partial cDNAs. Gene products were linked to enzymes reported in KEGG pathways (Release 32.0, October 1,

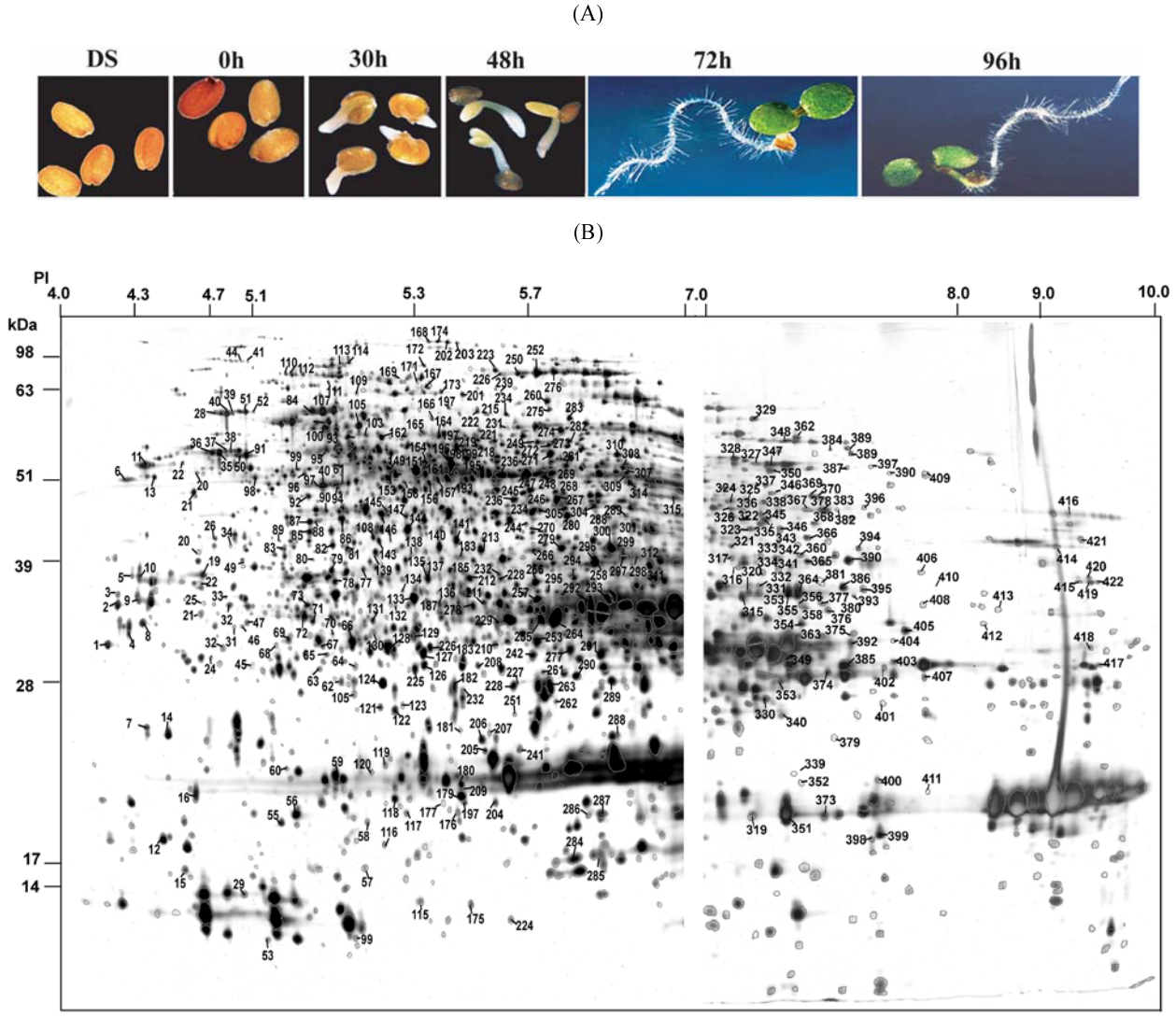

Fig. 1. Phenotypic characterization and silver-stained 2-DE analyses during Arabidopsis seed germination. A, Photos taken from various Arabidopsis seeds and early seedlings. DS, dry seeds; 0h, seeds after 3 days of stratification at 4°C; 30, 48, 72 and 96h, seeds placed under germinating conditions for specified amount of time (h). B, 2-DE analyses using 200 μg of proteins extracted from Arabidopsis seeds under germination for 30 h separated on 12.5% SDS-PAGE with pH ranging from 4 to 7 (left) or from 3 to 10 with pH 7 to 10 shown (right). All 2-DEs were performed in triplicates from independent protein samples.

2004) (Kanehisa *et al.*, 2004) based on enzyme commission (EC) numbers and were categorized according to Gene Ontology (GO, December 13, 2004) (The Gene Ontology Consortium, 2000).

Results and Discussions

Phenotypic and proteomic comparisons during seed germination and early seedling establishment Protein samples were prepared from dry and mature Arabidopsis seeds (Dry seeds), from seeds that were stratified at 4°C in ddH_2O for 3 days (0 h) plus being placed under germination conditions for 30 h, 48 h, 72 h and 96 h. The phenotypes of these seeds and seedlings were shown in Fig. 1A. We synchronized seedling growth and protein preparation simply by throwing out seeds that showed no visible radicles 30 h after beginning of the germination experiments. Systematic 2-DE gel analyses using quantified total protein samples prepared from these Arabidopsis materials were carried out on two different gels with pH ranging from 4-7 and 3-10. Representative gels using proteins prepared from 30 h germinating seeds were shown in Fig. 1B with acidic pH on the left and basic pH on the right side. Altogether, 1133 ± 5 protein spots were detected on the gel prepared with dry seeds while 1202 ± 2, 1167 ± 27, 1256 ± 15, 1315 ± 19 and 1309 ± 14 proteins spots were recovered from gels containing materials of 0, 30, 48, 72 and 96 h after germination initiation, respectively (data not shown). Seeds were stratified before initiation of the experiments to achieve uniform and also a higher rate of germination. No significant differences were

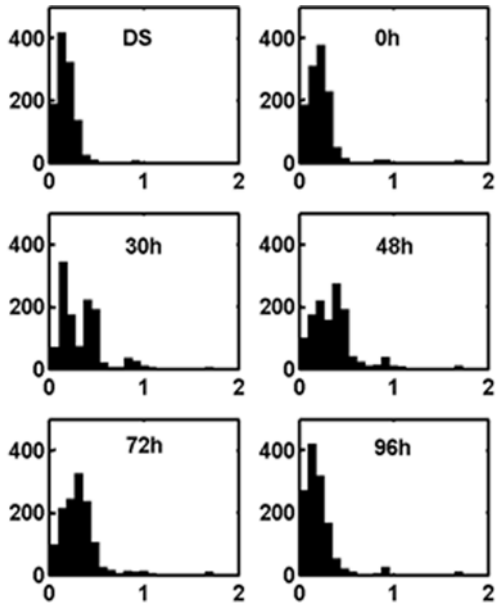

Fig. 2. Analysis of coefficient of variations for all replicating gels prepared from Arabidopsis materials harvested at the same stage. Y-axis denotes number of protein spots and X-axis denotes CV values.

observed on 2-DE gels prepared from germinating seeds with or without being treated at 4°C (data not shown). The same conclusion could be drawn when comparing our gels to those reported in Fig. 1C of ref (Gallardo *et al.*, 2001) as over 90% of the spots were matched. Seeds without being treated at 4°C for 3 d germinated about 20 h later than their cold-treated counterparts. Cruciferins, or 12S storage proteins (most major spots shown in Fig. 1B that were not numbered) were not included in the current manuscript since complex breakdown patterns were observed for these proteins. Coefficient of variations (CV) for each gel set representing a specific time point were calculated and reported in Fig. 2. CVs for more than 90% of the protein spots were smaller than 0.5 indicating that gels were reproducible. Quantitative analyses of all protein spots present on the gel of a given growth stage may provide a platform for further characterization of proteins involved in germination.

Identification of 437 proteins from Arabidopsis seeds and early seedlings Parallel to silver staining, replicates of the same 2-DE gels were stained with CBR350 and protein spots with signal intensities significantly above back ground levels (about 70% of the spots detected by silver staining) were excised, digested with trypsin and were analyzed by MALDI-TOF or TOF-TOF mass spectrometry. Of all these protein spots, only 637 produced mass peaks with 437 of them matched to peptides available in NCBInr with varying degrees of certainty. The molecular identities and their relative contents on respective gels were quantified in Supplementary Table with pattern analyses shown in Table 1. Over 67% of the identified proteins (293/437) were found at all stages

Table 1. Summary of protein expression patterns observed during Arabidopsis seed germination

Pattern[a]	Total number of proteins	↗	↘	→
111111	293	132	37	124
111110	11	0	4	7
111100	10	0	9	1
111000	3	0	3	0
110000	3	0	0	3
011111	31	16	1	14
001111	64	37	0	27
000111	11	8	0	3
000011	7	4	0	3
101111	2	0	2	0
101100	1	0	1	0
011100	1	0	1	0
Total	437	197	58	182

[a]A 6-dimension vector composed of 0, 1 was used to denote the presence (1) or absence (0) of a particular protein at the given stage. Every dimension of the vector was denoted a time point from left to right: dry seeds, 0 h, 30 h, 48 h, 72 h and 96 h.
[b]A line was drawn to fit the values of all time points relevant for the particular protein and the slope of that line was used to evaluate changes in protein contents. Proteins with slope rates >0.01 were considered as increasing (↗), those with slope rates <0.01 were considered as decreasing (↘) and those with slope rates fell in between were considered as unchanged (→), arbitrarily.

throughout our experiments. Among the sum, 124 kept stable levels while 132 increased gradually and 37 decreased gradually over all experimental stages. We suggest that these are housekeeping proteins needed for daily functions of the plant. As expected, a large number of proteins accumulated specifically during various stages of germination with 31 appeared at 0 h, 64 at 30 h, 11 at 48 h, 7 at 72 h (Tables 1 and 2). No new proteins appeared on gels containing samples 96h after germination. These results, together with phenotypic analysis shown in Fig. 1A, indicate actual completion of early seedling establishment around 72-96 h. Altogether, 27 proteins (excluding storage proteins) disappeared from various gels (Table 3). We suspected that different forms of proteins might be used for the same function in different growth stages of higher plants since many polypeptides in this group, such as peptidyl-prolyl cis-trans isomerase, mitochondrial import inner membrane (MIIM) translocase, ferritin, vacuolar ATP synthase, eukaryotic translation initiation factor (eTIF) and tryptophan synthase-related protein, were apparently very important during normal plant growth and development. Discrepancies in appearance or absence throughout the whole experimental period were observed for four protein spots with 2 defined as 101111, 1 each as 101100 and 011100 (Table 1). They might be regarded as experimental errors or might reflect true changes during Arabidopsis seed germination.

Table 2. Proteins that were significantly accumulated during Arabidopsis seed germination

Spot No.	PI/MW(kDa) [a]	NMP [b]	Locus	Annotation
011111				
2	4.2/32.7	6	At4g10480	nascent polypeptide associated complex alpha chain protein
24	4.9/28.7	6	At1g53850	20S proteasome alpha subunit E1
28	4.9/64.4	9	At4g24280	heat shock protein 70
46	5.1/30.7	11	At5g03630	monodehydroascorbate reductase
56	5.2/21.3	5	At2g31570	glutathione peroxidase
64	5.2/28.0	1	At1g21720	20S proteasome β subunit C1
65	5.2/29.1	5	At4g18040	eukaryotic translation initiation factor
88	5.2/44.4	12	At3g09820	adenosine kinase 1 (ADK1)
99	5.2/51.2	6	At4g32720	putative protein
118	5.3/23.6	9	At1g17290	alanine aminotransferase
132	5.3/32.4	5	At1g26630	eukaryotic translation initiation factor 5A
153 [c]	5.3/50.3	6	At1g49240	putative actin 8 protein
		6	At3g19450	Cinnamyl alcohol dehydrogenase
185	5.4/34.8	8	At1g23740	oxidoreductase, zinc-binding dehydrogenase family protein
192	5.4/49.2	12	At1g63940	monodehydroascorbate reductase
193	5.4/50.7	14	At3g17390	S-adenosylmethionine synthetase
204	5.5/23.4	1	At5g59880	actin-depolymerizing factor 3
214	5.5/42.0	4	At2g12405	hypothetical protein
236	5.6/50.8	18	AtCg00490	Ribulose bisphosphate carboxylase large chain precursor
239	5.6/94.5	7	At2g45030	mitochondrial elongation factor
258 [c]	5.8/36.5	7	At2g05990	enoyl-ACP reductase
		8	At3g44310	Nitrilase 1
285	6.1/17.6	6	At4g36440	expressed protein
293	6.1/32.7	6	At1g53240	malate dehydrogenase [NAD], mitochondrial
301	6.1/44.5	5	At4g39330	mannitol dehydrogenase
304	6.1/47.1	6	At5g54900	RNA-binding protein 45 (RBP45)
326	6.4/52.1	15	At1g14830	dynamin-like protein C (DL1C) nearly
331	6.5/37.9	10	At1g12900	glyceraldehyde 3-phosphate dehydrogenase, chloroplast
381	7.1/40.0	9	At2g40660	tRNA-binding region domain-containing protein
393	7.3/37.5	12	At3g46440	NAD-dependent epimerase/dehydratase family protein
400	7.5/26.5	9	At4g38740	peptidyl-prolyl cis-trans isomerase
001111				
3	4.2/34.0	7	At4g10480	nascent polypeptide associated complex alpha chain protein
6	4.3/50.6	15	At1g09210	calreticulin 2 (CRT2)
20	4.8/38.3	9	At5g09650	inorganic pyrophosphatase family protein
22	4.8/51.1	6	At5g39570	expressed protein
23	4.8/51.2	9	At5g39570	expressed protein
31	5.0/30.7	15	At1g74100	sulfotransferase family protein
33	5.0/33.5	12	At5g66570	oxygen-evolving enhancer protein 1-1, chloroplast
34	5.0/40.6	7	At1g62380	1-aminocyclopropane-1-carboxylate oxidase
35	5.0/52.3	13	At1g77510	protein disulfide isomerase
42	5.0/79.2	6	At3g54960	thioredoxin family protein
43	5.0/79.3	12	At3g54960	thioredoxin family protein
44	5.0/79.4	12	At2g04030	heat shock protein
48	5.1/32.5	11	At3g53460	29 kDa ribonucleoprotein, chloroplast
52	5.1/65.0	9	At5g49910	heat shock protein 70
54	5.2/5.0	9	At4g14980	CHP-rich zinc finger protein
67	5.2/29.3	8	At5g16710	dehydroascorbate reductase
68	5.2/29.4	7	At3g50820	oxygen-evolving enhancer protein, chloroplast
70	5.2/29.8	6	At1g15140	oxidoreductase NAD-binding domain-containing protein
75	5.2/33.5	6	At5g65690	phosphoenolpyruvate carboxykinase
78	5.2/34.8	7	At3g10670	ABC transporter family protein
79	5.2/36.3	9	At1g05010	1-aminocyclopropane-1-carboxylate oxidase
81	5.2/39.2	7	At1g70830	Bet v I allergen family protein

Table 2. Continued

Spot No.	PI/MW(kDa) [a]	NMP [b]	Locus	Annotation
85	5.2/42.5	10	At5g63980	3'(2'),5'-bisphosphate nucleotidase
91	5.2/46.4	16	At2g28000	RuBisCO subunit binding-protein alpha subunit, chloroplast
97	5.2/49.8	13	At3g13470	chaperonin
98	5.2/50.2	12	At5g44340	tubulin beta-4 chain (TUB4) nearly
128	5.3/29.0	6	At3g46940	deoxyuridine 5'-triphosphate nucleotidohydrolase family
130[C]	5.3/29.9	7	At3g54640	tryptophan synthase, alpha subunit
		12	At5g14740	Carbonic anhydrase 2
137	5.3/36.3	10	At5g66190	ferredoxin--NADP(+) reductase
142	5.3/40.3	6	At3g02780	isopentenyl-diphosphate delta-isomerase II
165	5.3/55.3	11	At3g08590	2,3-biphosphoglycerate-independent phosphoglycerate mutase
167	5.3/64.2	6	At5g48300	glucose-1-phosphate adenylyltransferase small subunit 1
168	5.3/77.5	10	At1g50380	prolyl oligopeptidase family protein
181	5.4/26.4	7	At5g28050	cytidine/deoxycytidylate deaminase family protein
183	5.4/28.5	7	At3g62030	peptidyl-prolyl cis-trans isomerase, chloroplast
207	5.5/26.3	7	At1g02560	ATP-dependent Clp protease proteolytic subunit (ClpP1)
216	5.5/46.9	9	At4g01850	S-adenosylmethionine synthetase 2 (SAM2)
217	5.5/49.7	8	At3g17390	S-adenosylmethionine synthetase
278	6.0/28.9	6	At1g07890	L-ascorbate peroxidase 1, cytosolic
281	6.0/52.8	11	At1g12000	pyrophosphate--fructose-6-phosphate 1-phosphotransferase beta subunit
306	6.1/47.8	13	At1g77120	alcohol dehydrogenase (ADH)
316	6.3/42.4	12	At4g35200	hypothetical protein
344	6.6/48.7	7	At2g38700	mevalonate diphosphate decarboxylase
345	6.6/50.5	8	At2g39770	GDP-mannose pyrophosphorylase
350	6.7/50.9	13	At4g13930	glycine hydroxymethyltransferase
353	6.8/29.1	7	At3g22110	20S proteasome α subunit C (PAC1)
358	6.8/36.3	11	At2g46280	eukaryotic translation initiation factor 3 subunit 2
359	6.8/38.3	10	At4g05390	ferredoxin--NADP(+) reductase
361	6.8/50.6	18	AtCg00490	hypothetical protein
370[C]	6.9/50.9	13	At1g52100	hypothetical protein
		26	At4g37930	glycinehydroxymethyl transferase
371	6.9/55.7	12	At1g49760	polyadenylate-binding protein
372	6.9/62.6	12	At4g37870	phosphoenolpyruvate carboxykinase
375	7.0/32.6	6	At5g66140	20S proteasome α subunit D1 (PAD1)
377	7.0/36.4	8	At1g30510	ferredoxin--NADP(+) reductase
379	7.1/27.4	11	At1g56450	20S proteasome â subunit G1 (PBG1)
384	7.1/57.7	12	At1g78570	NAD-dependent epimerase/dehydratase family protein
385	7.2/30.1	12	At1g59900	pyruvate dehydrogenase E1 compoent alpha subunit, mitochondrial
391	7.2/58.2	13	At1g78570	NAD-dependent epimerase/dehydratase family protein
404	7.7/32.2	7	At5g15090	porin
416	9.5/48.9	11	At4g17520	nuclear RNA-binding protein
417	9.6/30.2	11	At2g33150	acetyl-CoA C-acyltransferase
418	9.6/31.6	17	At1g74470	geranylgeranyl reductase
000111				
41	5.0/78.7	14	At2g04030	heat shock protein
186	5.4/36.8	11	At2g39990	26S proteasome regulatory subunit
288	6.1/26.2	9	At4g14800	20S proteasome β subunit D2 (PBD2)
317	6.3/44.2	10	At4g35630	phosphoserine aminotransferase, chloroplast (PSAT)
336	6.5/49.9	13	At1g43190	polypyrimidine tract-binding protein
349	6.7/31.0	12	At4g35000	L-ascorbateperoxidase
351	6.8/25.7	8	At5g08280	hydroxymethylbilane synthase
376	7.0/33.0	8	At5g66510	bacterial transferase hexapeptide repeat-containing protein
390	7.2/58.1	15	At5g04590	sulfite reductase
410	8.1/35.4	14	At1g18080	WD-40 repeat family protein
413	8.8/3.8	10	At5g15090	porin

Table 2. Continued

Spot No.	PI/MW(kDa) [a]	NMP[b]	Locus	Annotation
000011				
356	6.8/34.0	14	At3g14220	GDSL-motif lipase
365	6.9/43.5	10	At1g48850	chorismate synthase
378	7.0/50.5	8	At4g35090	catalase 2
380	7.1/35.7	7	At4g37800	xyloglucan:xyloglucosyl transferase
388	7.2/51.8	15	At1g20620	catalase 3 (SEN2)
397	7.4/51.9	14	At1g20620	catalase 3 (SEN2)
409	8.0/51.0	11	At5g26780	glycine hydroxymethyltransferase

[a]Experimental PI and MW.
[b]Number of matched peptides.
[c]Spots produced two different proteins. See Table 5 for details.

Table 3. Proteins that disappeared during different stages of Arabidopsis seed germination

Spot No.	PI/MW(kDa)[a]	NMP[b]	Locus	Annotation
110000				
401	7.6/28.4	7	At4g16160	mitochondrial import inner membrane translocase subunit
407	8.0/29.6	8	At4g38740	peptidyl-prolyl cis-trans isomerase
411	8.2/23.5	6	At3g46110	expressed protein
111000				
61	5.2/27.0	8	At3g11050	ferritin
241	5.7/25.9	4	At3g51810	Em-like protein GEA1 (EM1)
309	6.1/50.6	7	At3g21380	jacalin lectin family protein
111100				
90	5.2/46.1	32	At1g76030	vacuolar ATP synthase subunit B
175	5.4/8.7	4	At1g27970	nuclear transport factor 2, putative
177	5.4/22.8	7	At5g58070	lipocalin
246	5.7/49.3	13	At1g63940	monodehydroascorbate reductase
251	5.8/27.5	25	At5g35820	copia-like retrotransposable element
274	5.9/57.4	18	At2g42560	late embryogenesis abundant domain-containing protein
398	7.5/24.2	6	At1g14930	dynamin-like protein C (DL1C) nearly
399	7.5/24.5	8	At1g14940	major latex protein-related / MLP-related low
402	7.6/30.2	9	At4g16160	mitochondrial import inner membrane translocase subunit
403	7.7/30.6	5	At5g35040	hypothetical protein
111110				
53	5.2/3.7	8	At3g07780	expressed protein
69	5.2/29.6	11	At5g16440	isopentenyl-diphosphate delta-isomerase I
72	5.2/31.5	10	At1g11910	aspartic proteinase family protein
74	5.2/33.3	12	At1g11840	lactoylglutathione lyase
166	5.3/60.6	10	At5g44320	eukaryotic translation initiation factor 3 subunit 7
191	5.4/49.1	9	At2g33070	jacalin lectin family protein
201	5.4/78.0	11	At3g13460	expressed protein
247	5.7/49.4	12	At4g02930	elongation factor Tu
323	6.4/48.7	16	At5g38530	tryptophan synthase-related
342	6.6/44.2	7	At5g63620	oxidoreductase, zinc-binding dehydrogenase family protein
396	7.4/50.7	9	At1g20630	catalase 1

[a]Experimental PI and MW.
[b]Number of matched peptides.

Several transcription factor-like proteins, such as the eTIF encoded separately by At1g26630, At4g18040 and At2g46280, the nascent polypeptide-associated complex alpha chain protein encoded by At4g10480, the CHP-rich zinc finger protein encoded by At4g14980 were found to be synthesized specifically before or at the time of radicle emergence (Table 2). Quite a few potentially important proteins like the adenosine kinase (ADK) encoded by At3g09820, the ferredoxin-NADP(+) reductases encoded separately by At5g66190, At1g30510 and At4g05390, the phosphoenolpyruvate carboxykinase (PEPCK)

Table 4. GO classification of all proteins identified currently

Function	No. of independent genes
Enzymes	184
Binding proteins	75
Transporters	20
TF & chaperones	14
Structural proteins	9
Antioxidant proteins	3
Unknown proteins	50
Total	355

encoded by At4g37870 and porin encoded by At5g15090 were found to be *de novo* synthesized early in seedling development (Table 2). ADK were proposed to be components of ethylene and phytochrome-mediated light signal transduction pathways (Novikova *et al*., 2003; Im *et al*., 2004) while PEPCK were required for photosynthetic functions (Lebreton *et al*., 2003). In rice, the nascent polypeptide-associated complex alpha chain protein was found to accumulate significantly after salt stress (Yan *et al*., 2005). Quite a number of enzymes known to participate in gibberellin or ethylene biosynthetic pathways, including the geranylgeranyl reductase encoded by At1g74470, the mevalonate diphosphate decarboxylase (MDD) encoded by At2g38700, two 1-aminocyclopropane-1-carboxylate oxidases (ACO) encoded by At1g05010 and At1g62380, two S-adenosylmethionine synthetases encoded by At3g17390 and At4g01850, also appeared at the same growth stages (Table 2).

A JUMONJI family transcription factor (protein spots number 55 encoded by At4g20400, Supplementary Table) showed highest amount in dry seeds with low but persistent levels seen in all other time points. Members of this family were implicated in animal or human cell growth and heart development (Kim *et al*., 2003; Schofield and Ratcliffe, 2004), or were involved in regulating flowering time in *Arabidopsis thaliana* (Noh *et al*., 2004).

Gene identification and localization in various metabolic pathways We next sorted all protein-coding genes manually and found that 355 independent genes were involved in producing the identified proteome. They were categorized by GO and were reported in Table 4. Most of them were enzymes and coenzymes (binding proteins as defined by GO). Of the total number of genes identified, 226 were successfully linked to proteins in different metabolic pathways as characterized in KEGG database. Proteins encoded by these genes were quantified according to their normalized signal intensities in each gel and were reported in Fig. 3. We found that polypeptides involved in amino acids, carbohydrates and energy metabolism consisted for about 1/4 of the total proteome with another 3% involved in GIP on gels containing proteins samples harvested 72h after germination initiation (Fig. 3). Throughout the whole experimental period, the

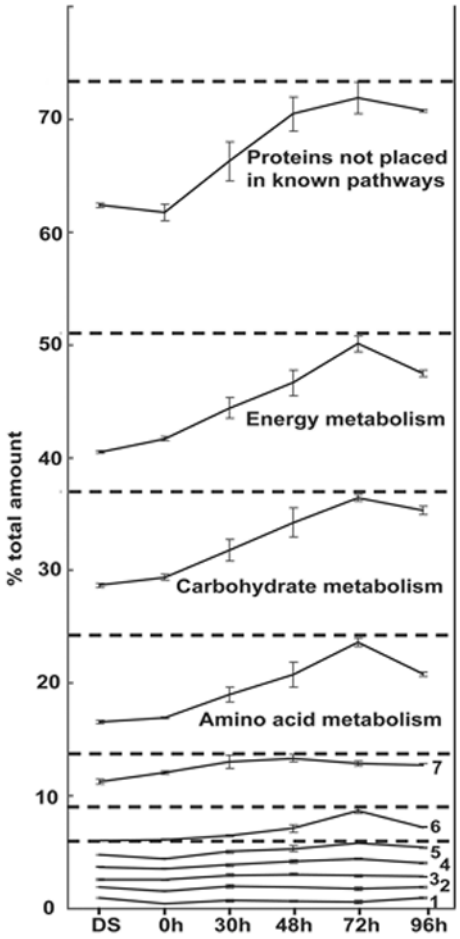

Fig. 3. Several major categories increased in relative levels rapidly over the course of germination until around 72h. Normalized signal intensities of all protein spots in a given category from each gel were mathematically added up and were analyzed throughout the whole experimental period. Since quite many genes showed redundant functions in GO, the grand total of this figure was about 125-130%. Categories numbered 1 to 7 are: transporters, signal transduction, nucleotide metabolism, lipid metabolism, biosynthesis of glycan and secondary metabolites, metabolism of cofactors and vitamins, genetic information process, respectively.

relative amount of proteins related to metabolism of cofactors and vitamins (MCV), to biosynthesis of glycan and secondary metabolites (BGSM), and to GIP increased very sharply. The relative amount of proteins categorized as transporters together with those in signal transduction was high in dry seeds, decreased to low levels after stratification and increased soon after seed germination (Fig. 3). All 8 enzymes in glycolytic pathway from D-Xylulose-5P to pyruvate production (Fig. 4A), 5 enzymes in cabon fixation pathway from ribose-5P to glyceraldehyde-3P (Fig. 4B) and several enzymes that reduce ammonia to glutamine, pyruvate and asparagines (Fig. 4C) were identified and quantified. The intensities for most of these enzymes increased significantly over the experimental period (Fig. 4).

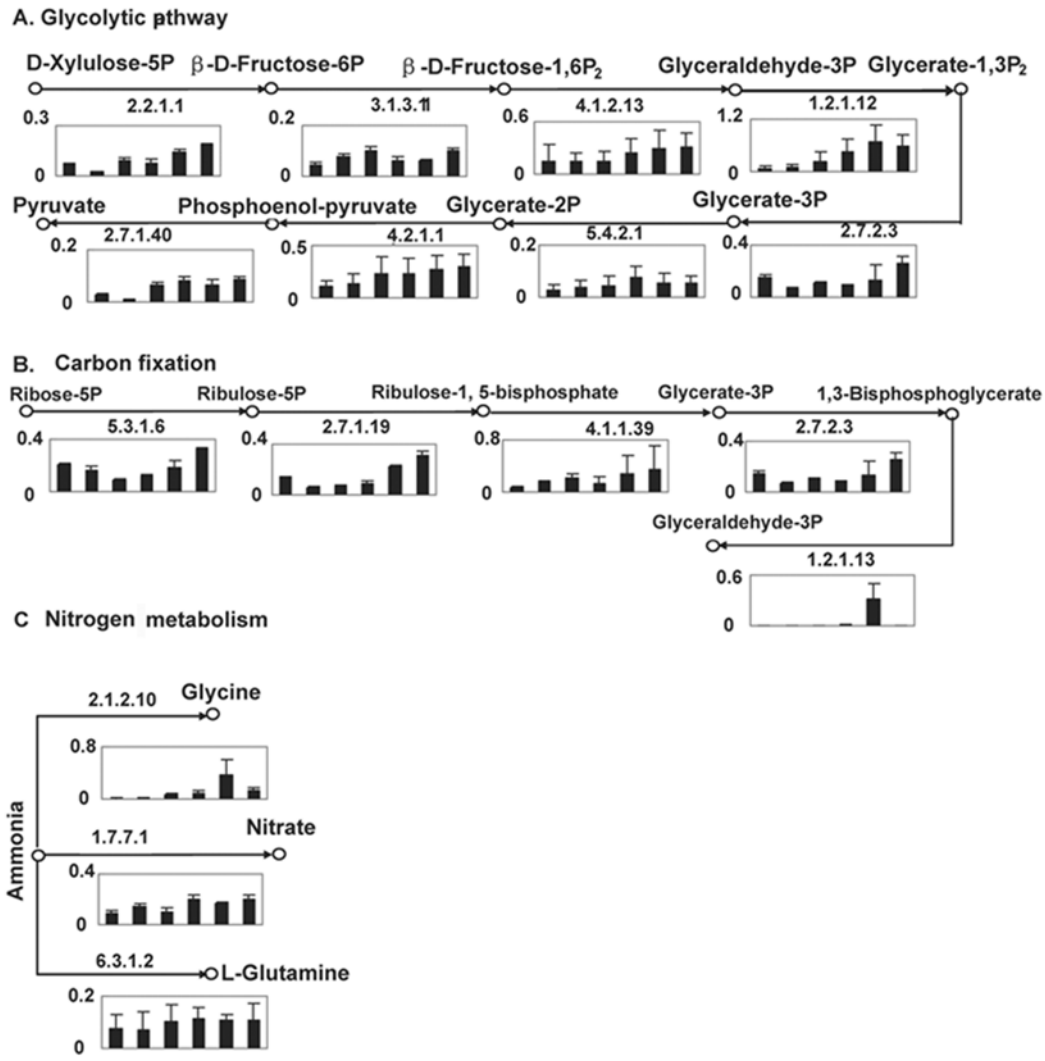

Fig. 4. Quantitative analyses of enzymes involved in: A, the glycolytic pathway from D-xylulose-5P to pyruvate; B, in carbon fixation pathway from ribose-5P to glyceraldehyde-3P; C, in ammonia metabolism that produces Nitrite, L-asparagine as well as L-glutamine. 2.2.1.1, transketolase; 3.1.3.11, fructose-1, 6-bisphosphatase; 4.1.2.13, fructose-bisphosphate aldolase; 1.2.1.12, glyceraldehyde 3-phosphate dehydrogenase; 2.7.2.3, phosphoglycerate kinase; 5.4.2.1, phosphoglycerate mutase; 4.2.1.11, enolase; 2.7.1.40, pyruvate kinase. 5.3.1.6, ribose 5-phosphate isomerase; 2.7.1.19, phosphoribulokinase; 4.1.1.39, ribulose-bisphosphate carboxylase; 2.7.2.3, phosphoglycerate kinase; 1.2.1.13, glyceraldehyde-3-phosphate dehydrogenase. 3.5.1.1, L-asparaginase; 1.7.7.1, ferredoxin-nitrite reductase; 6.3.1.2, glutamine synthetase.

Identification and analyses of putative protein complexes purified from germinating Arabidopsis seeds or early seedlings During the protein identification process, a small, but significant number of spots perplexed us by producing peptide maps from two entirely different proteins. As reported in Table 5, 30 polypeptides were produced from 15 gel spots with each contained two identifiable proteins. Among all these partners, interactions between betaine-aldehyde dehydrogenase (BAD, encoded by At1g74920) and the dihydrolipoamide S-acetyltransferase (DSA, encoded by At3g13930) were predicted by the presence of Aldedh and Biotin_lipoyl domains respectively on one of the proteins (Lebreton *et al*., 2003). Another pair of proteins, the protochlorophyllide reductase B (encoded by At4g27440) and ribosomal protein L1 family protein (encoded by At3g63490), possessed functional domains similar to that predicted in the database that were likely to interact (Ng *et al*., 2003). Further, one phosphoglycerate/bisphosphoglycerate mutase family protein (At3g26780) and one ketol-acid reductoisomerase (At3g58610) were found to probably team up with chaperonin proteins encoded by At5g16070 or At5g20890, respectively (Table 5). We suggested that the presence of a chaperonin protein might actually help stabilize the enzymes. However, most of the proteins recovered from complexes showed very different theoretical MW and PI, and should not present on the gel as a single spot. We have no explanation as to why they came out together on the gel. We suggest that they might reflect an *in vivo* interaction between the two partners for some unknown

Table 5. Proteins obtained from putative complexes

Spot No	PI/MW(kDa)[a]	NMP[b]	Locus	Annotation
17	4.8/32.0	9	At1g35160	14-3-3 protein homolog GF14
		7	At4g22240	Fibrillin-like protein
26	4.9/42.0	10	At2g31670	Expressed protein
		12	At4g15510	photosystem II reaction center PsbP family protein
73	5.2/32.6	13	At1g11910	aspartyl protease family protein
		7	At5g07460	peptide methionine sulfoxide reductase
102	5.2/53.5	8	At1g74920	betaine-aldehydedehydrogenase
		11	At3g13930	dihydrolipoamideS-cetyltransferase
130	5.3/29.9	7	At3g54640	tryptophan synthase, alpha subunit
		12	At5g14740	Carbonic anhydrase 2
153	5.3/50.3	6	At1g49240	putative actin 8 protein
		6	At3g19450	Cinnamyl alcohol dehydrogenase
194	5.4/51.0	7	At3g08590	2,3-biphosphoglycerate-independent phosphoglycerate mutase
		13	At5g66760	succinate dehydrogenase
188	5.4/38.5	7	At2g47470	Thioredoxin family protein
		8	At3g44320	Nitrilase 3
197	5.4/52.1	11	At3g58610	ketol-acid reductoisomerase
		20	At5g20890	chaperonin
233	5.6/48.4	12	At2g27860	hypothetical protein
		10	At1g02500	S-adenosylmethionine synthetase 1
258	5.8/36.5	7	At2g05990	enoyl-ACP reductase
		8	At3g44310	Nitrilase 1
273	5.9/53.5	4	At3g26780	bisphosphoglycerate mutase family protein
		7	At5g16070	TCP-1 chaperonin-like protein
343	6.6/46.3	6	At2g30970	aspartate aminotransferase
		6	At1g54010	myrosinase-associated protein
370	6.9/50.9	13	At1g52100	hypothetical protein
		26	At4g37930	glycinehydroxymethyl transferase
422	9.7/39.7	8	At4g27440	protochlorophyllide reductase precursor
		10	At3g63490	Ribosomal protein L1p

[a]Experimental PI and MW.
[b]Number of matched peptides.

biological functions or they might be considered as artifacts produced during protein purification process, or both. Our data represent a major step toward identifying full proteomic changes during Arabidopsis seed germination.

Acknowledgments We thank Dr. James Brush for help with setting up the experiments. We are grateful to Jin-Dong Zhao for critical reading of the manuscript. This work was supported by a grant from China National Natural Science Foundation (Grant 30221120261) and the China National Basic Research Program (Grant 2004CB117302).

References

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796-815.

Bewley, J. D. (1997) Seed germination and dormancy. *Plant Cell* **9**, 1055-1066.

Bove, J., Jullien, M. and Grappin, P. (2001) Functional genomics in the study of seed germination. *Genome Biol.* **3**, 1002.1-1002.5

Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2001) Proteomic analysis of arabidopsis seed germination and priming. *Plant Physiol.* **126**, 835-848.

Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2002) Proteomics of Arabidopsis seed germination. a comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* **129**, 823-837.

Gamble, S. C., Dunn, M. J., Wheeler, C. H., Joiner, M. C., Adu-Poku, A. and Arrand, J. E. (2000) Expression of proteins coincident with inducible radioprotection in human lung epithelial cells. *Cancer. Res.* **60**, 2146-2151.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol.* **124**, 1570-1581.

Gorg, A., Postel, W. and Gunther, S. (1988) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects. *Electrophoresis* **9**, 531-546.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, 277-280.

Koornneef, M., Bentsink, L. and Hilhorst, H. (2002) Seed dormancy and germination. *Current Opin. Plant Biol.* **5**, 33-36.

Koornneef, M. and Karssen, C. M. (1994) Seed dormancy and germination; in *Arabidopsis*, Meyerowitz, E. M. and Somerville, C. R. (eds.), pp. 313-334, Cold Spring Harbor Laboratory Press, New York, USA.

Kim, T-G., Kraus, J. C., Chen, J. and Lee, Y. (2003) JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. *J. Biol. Chem.* **278**, 42247-42255.

Lebreton, S., Graciet, E. and Gontero, B. (2003) Modulation, via protein-protein interactions, of glyceraldehyde-3-phosphate dehydrogenase activity through redox phosphoribulokinase regulation. *J. Biol. Chem.* **278**, 12078-12084.

Im, Y. J., Kim, J. I., Shen, Y., Na, Y., Han, Y. J., Kim, S. H., Song, P. S. and Eom, S. H. (2004) Structural analysis of arabidopsis thaliana nucleoside diphosphate kinase-2 for phytochrome-mediated light signaling. *J. Mol. Biol.* **343**, 659-670.

Ng, S. K., Zhang, Z., Tan, S. H. and Lin, K. (2003) InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res.* **31**, 251-254.

Noh, B., Lee, S. H., Kim, H. J., Yi, G., Shin, E. A., Lee, M., Jung, K. J., Doyle, M. R., Amasino, R. M. and Noh, Y. S. (2004) Divergent roles of a pair of homologous jumonji/zinc-finger–class transcription factor proteins in the regulation of arabidopsis flowering time. *Plant Cell* **16**, 2601-2613.

Novikova, G. V., Moshkov, I. E., Smith, A. R. and Hall, M. A. (2003) Nucleoside diphosphate kinase is a possible component of the ethylene signal transduction pathway. *Biochemistry (Moscow)* **68**, 1342-1348.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C. and Job, D. (2004) The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mrnas during germination. *Plant Physiol.* **134**, 1598-1613.

Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L. A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* **31**, 224-228.

Schofield, C. J. and Ratcliffe, P. J. (2004) Oxygen sensing by HIF hydroxylases. *Nat. Rev. Mol. Cell Biol.* **5**, 343-354.

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

The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29.

Tyler, L., Thomas, S. G., Hu, J., Dill, A., Alonso, J. M., Ecker, J. R. and Sun, T. P. (2004) DELLA proteins and gibberellin-regulated seed germination and floral development in arabidopsis. *Plant Physiol.* **135**, 1008-1019.

Yamauchi, Y., Ogawa, M., Kuwahara, A., Hanada, A., Kamiya, Y. and Yamaguchi, S. (2004) Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds. *Plant Cell* **16**, 367-378.

Yan, S., Tang, Z., Su, W. and Sun, W. (2005) Proteomic analysis of salt stress-responsive proteins in rice root. *Proteomics* **5**, 235-244.