Separation and Characterization of Spikelet Proteins at Young Microspore Stage in Rice

Sun Hee Woo*†, Tae Seon Kim*, Seong Woo Cho*, Jung Gu Ahn*, Keun Yook Chung, Byung-Moo Lee***, Yong Gu Cho*, Hong Sig Kim*, Beom Heon Song*, Chul Won Lee*, and Seung Keun Jong***

**Dept. of Crop Science, Chungbuk National University, Cheongju 361-763, KOREA*
***Dept. of Agriculture Chemistry, Chungbuk National University, Cheongju 361-763, KOREA*
****Dept. of Plant Biotechnology, Dongguk University, Seoul 100-715, KOREA*

ABSTRACT : Spikelet proteins expressed at the young microspore stage in rice were separated and analysed by two-dimensional polyacrylamide gel electrophoresis (2-DE). The separated proteins were electroblotted onto a polyvinylidene difluoride (PVDF) membrane, and 50 proteins were analyzed by a gas-phase protein sequencer. The *N*-terminal amino acid sequences of 20 out of 50 proteins were determined. *N*-terminal regions of the remaining proteins could not be sequenced because of blocking. The internal amino acid sequences of proteins were determined by sequence analysis of peptides obtained by the Cleveland peptide mapping method. Results revealed the presence of the photosynthetic apparatus at rice young microspore stage. Major proteins identified in this study could be used as a marker for various studies on physiological stresses.

Keywords: rice, young microspore stage, two-dimensional gel, cleveland peptide mapping, amino acid sequence analysis, proteome analysis, EST

Rice is a staple food for about a half of people in the world and is mainly harvested and consumed in Asia and Africa, where the population is be expected to double during the next 50 years. However, increase of the rice production is becoming more difficult because of various biotic and abiotic stresses. To meet the requirements of incoming people, it is necessary to develop new high yielding rice cultivars with high protein content. High protein content is particularly important for the people of South and Southeast Asia as their daily food consumption is dominated by rice (Islam *et al.*, 1996). To achieve the current breeding goals for high yield and high protein rice, innovative tools are essential to identify the huge informations hidden in the rice proteome, which controls all characteristics of rice plants.

Several methods, including serial analysis of gene expression, oligonucleotide and cDNA microarrays, and large-scale sequencing of expressed sequence tags have been developed to measure globally and quantitatively gene expression at the mRNA level (Velculescu *et al.*, 1995). The discovery of post-transcriptional mechanisms that control rate of synthesis and half-life of proteins (Varshavsky, 1996) and the ensuing nonpredictive correlation between mRNA and protein levels expressed by a particular gene (Futcher *et al.*, 1999; Gygi *et al.*, 1999) indicated that direct measurement of protein expression is also essential for the analysis of biological precesses and systems.

Global analysis of gene expression at the protein level is commonly termed as proteomics. Proteome analysis is performed to make the best use of the data obtained in the genome research project according to the following steps; (1) A number of proteins are separated by mainly two-dimensional electrophoresis (2-DE), (2) Based on the peptide-mass fingerprint and partial amino acid sequence, the proteins separated by 2-DE are identified by the use of protein database, and simultaneously the genes encoding them are identified in the gene library constructed in the genome project, (3) The proteins without known function are characterized in amount, localization, structure, post-translational modification, enzyme activity and protein- protein interaction, etc., and (4) The functions of a number of proteins and genes are determined with the data obtained from both protein and gene analyses, which seems to be the most important and difficult step in proteome analysis.

Recently, the peptide-mass fingerprinting by matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF/ MS) is usually used for the identification of proteins and genes (Shevchenko et al. 1996). Although this is a high throughput method, it is not always useful to identify the proteins and genes in plants as well as the other eukaryotes. Because plant proteins are often post-translationally modified, it is difficult to identify the proteins and genes by the peptide-mass fingerprint. The amino acid sequence of, at least, a peptide consisting of 5 ~ 6 residues is required to identify the proteins and genes (Hirano, 1992). The mass-spectrometry such as electrospray ionization quadrupole time-of-flight (ESI Q-TOF/ MS) can be used to

†Corresponding author: (Phone) +82-43-261-2515 (E-mail) shwoo @chungbuk.ac.kr <Received February 3, 2006>

determine the partial amino acid sequences of proteins separated by 2-DE (Wilm *et al.*, 1996). This is a highly sensitive and accurate, but not a high-throughput method.

On the other hand, a gas-phase protein sequencing is neither rapid nor sensitive compared with mass spectrometry in identification of charateristics of proteins, but still useful to determine the partial amino acid sequences of proteins separated by 2-DE. Presently, subpicomole amounts of proteins separated by 2-DE and electroblotted onto a PVDF membrane can be sequenced by the gas-phase sequencer. Based on the sequences obtained, we can precisely identify the proteins and genes.

However, if the *N*-terminal amino groups of proteins are modified, we cannot determine the *N*-terminal amino acid sequences of the proteins by Edman degradation. There are a lot of proteins with blocking groups at the *N*-terminus in plants. In this case, the Cleveland peptide mapping (Cleveland *et al.*, 1977) and gas phase sequencing after separation by 2-DE to obtain the internal sequence data are useful (Hirano *et al.*, 1992). Separation of proteins by 2-DE, and electroelution from the gels were followed by dialyzation of the proteins and lyophilization. The resultant protein samples were subjected to the Cleveland peptide mapping. However, it takes long time to perform the electroelution, dialysis and lyophilization.

Here, we report a simple and relatively rapid sample preparation method for the Cleveland peptide mapping and its application to identify the young microspore stage proteins separated by 2-DE in the rice plants.

MATERIALS AND METHODS

Plant materials and sample preparation

Rice (cv. Nipponbare) were cultivated under controlled conditions (8 hour light intensity, 70% relative humidity and 20°C constant temperature). At the young microspore stage, young spikelets were collected 10 ~ 13 days after heading date.

Two-dimensional gel electrophoresis

A portion (500 mg) of the young spikelets was removed, homogenized with 1 ml of lysis buffer (O'Farrell, 1975) and centrifuged at 15000 × g for 5 minutes. The supernatant (50 μl) was used to 2-DE (Hirano, 1982). IEF was carried out in a glass capillary tube which is 13cm in length and 3 mm in diameter. The gel solution consisted of 10% NP-40, 30% w/v acrylamide, 9.5M urea, 10% ammonium persulfate, and the equal mixture of 2% carrier ampholytes (pH 3.5 ~ 10 and 5 ~ 8). The sample overlay buffer consisted of 20 μl of 1/2 lysis buffer. Electrophoresis was carried out at 200 V for 30 minutes, at 300 V for 1 hour, followed by 600 V for 16 hour and 800 V for 1 hour. Sodium dodecyl sulfate (SDS)-PAGE in the second dimension was performed with 15% separation gels and 5% stacking gels at a constant current of 30 mA. The isoelectric point and relative molecular weight of each protein were determined using molecular weight marker proteins (Amersham Pharmacia Biotech).

Image analysis of the electrophoresis pattern

After staining, the electrophoresis patterns were scanned using a flatbed scanner, and the data were analyzed using PDQuest software version 6.1 (Bio-Rad Laboratories, CA).

N-terminal sequence analysis

Following separation by 2-DE, proteins were electroblotted onto a PVDF membrane (ProBlot, Applied Biosystems, Foster City) and detected by Coomassie Brilliant Blue R250 staining (Hirano *et al.*, 1990). They were cut out of PVDF membranes and applied to the upper glass block of the reaction chamber in a gas-phase protein sequencer (491, Applied Biosystems). Edman degradation was performed according to the standard program supplied by Applied Biosystems. The released phenylthiohydantoin amino acids were identified by the on-line system of high performance liquid chromatography (120A, Applied Biosystems).

Internal sequence analysis

The proteins were separated by 2-DE and stained with Coomassie Brilliant Blue R250. After destaining, the gels were dried up and gel pieces containing protein spots were removed with a cutter. The gel pieces were chopped into smaller pieces and inserted in the sample well of the stacking gel for SDS-PAGE. One hundred ml of the electrode solution was added to the dried gel pieces. After incubation for 1 hour, 20 ml of 2x diluted SDS sample buffer containing 10 ml of *Staphylococcus aureus* V8 protease (Pierce, Rockford) (0.1 mg/ml) in deionized water was overlaid on the sample solution. Electrophoresis was performed until the sample and protease were stacked in the upper gel, and interrupted for 1 hour to digest the protein (Cleveland *et al.*, 1977). Electrophoresis was then continued and the separated digests were electroblotted on the PVDF membrane and subjected to gas-phase sequencing.

Homology search of amino acid sequence

The amino acid sequences obtained were compared with

those of proteins compiled in the protein sequence database (SWISS-PROT and NCBI) using FASTA (http://www2.ebi.ac.uk/fasta3), or BLAST (http://www.ncbi. nlm.nih.gov/blast) programs.

Identification of cDNA encoding the sequenced protein

The both cDNAs or ESTs sequences obtained in the Rice Genome Research Project (DDBJ (http://www.ddbj.nig.ac.jp) and NCBI (http://www.ncbi.nlm.nih.gov/) have been translated into protein sequences and compiled as a database. Using this database, the cDNAs encoding the proteins partially sequenced in the present study were screened according to the method described by Pearson and Lipman (1988).

RESULTS AND DISCUSSION

Since 1987, in Japan, an amino acid sequence data-file on rice proteins has been in the process of construction. This data file contains the partial amino acid sequences and other biochemical and physiological characteristics of the protein. A high-priority goal in the program has been to establish a database of partial sequences of all proteins interfacial with further DNA sequence information from the Rice Genome Research Project (Sasaki *et al.*, 1994).

In the present study, we focused on separation and identification of spikelet proteins expressed at the young microspore stage. More than 250 major spots were identified by CBB staining in 2-D PAGE patterns of the young microspore stages (Fig. 1).

Fifty protein spots were selected from the young microspore stage. Twenty out of 50 proteins were determined (Table 1). *N*-terminal regions of the remaining proteins could not be sequenced and they were inferred to have a blocking group at the *N*-terminus.

Among the sequenced 50 proteins, 18 were new proteins which listed in SWISS-PROT and NCBI databases. The relatively low identification ratio and higher percentage of new sequence in proteins of tobacco, wheat, barley and maize as

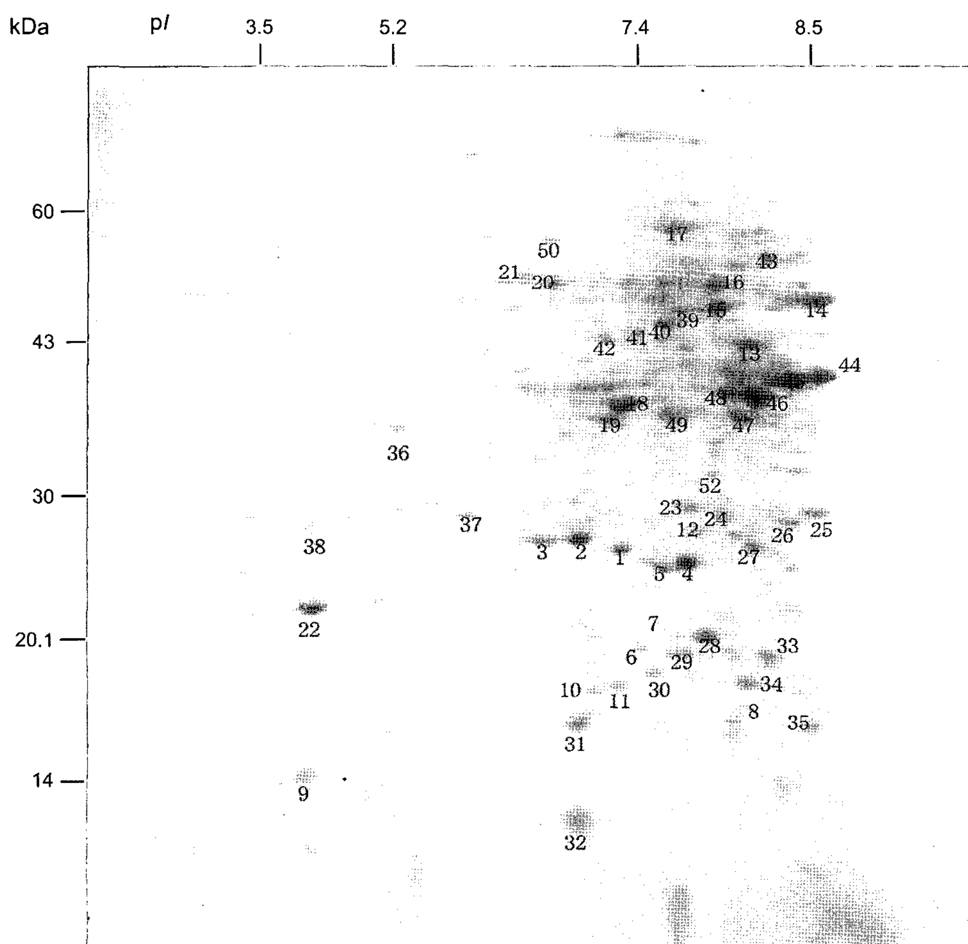

Fig. 1. Two-dimensional electrophoresis pattern of the rice spikelet proteins were separated by isoelectric focusing in the first dimension (right to left) and SDS-PAGE in the second dimension (up to down) and detected by Coomassie Blue staining.

Table 1. List of microsequenced rice spikelet proteins expressed at the young microspore stage.

Spots no.	Sequence[a)]	Homologous protein	Species	Accession no.	EST Acc. no.[c)]	pI	MW (kDa)
1	AYGEAANVFGKAK (13/13)	OXYGEN-EVOLVING ENHANCER PROTEIN 2 (27.3 kDa)	WHEAT	Q00434†	A1759921	7.1	27.4
2	AKNYPTVSAEYSEA (13/14)	UNKNOWN PROTEIN FROM 2D-PAGE OF ETIOLATED COLEOPTILE (19.4 kDa)	MAIZE	P80627†	BE512153	6.7	27.9
3	GRKFFVGGNWKCNGT (15/15)	TRIOSEPHOSPHATE ISOMERASE, CYTOSOLIC (27.1 kDa)	RICE	P48494†	BE607405	6.5	27.8
4	GLPAAIGAAVGRPD (14/14)	ACETOLACTASE SYNTHASE II. CHLOROPLAST PRECURSOR (72.4 kDa)	TOBACCO	P09114†	C72508	7.7	26.2
5	VTTVALPDLP (10/10)	SUPEROXIDE DISMUTASE, MITOCHONDRIAL PRECURSOR (24.9 kDa)	RICE		C20060	7.2	25.8
6	GPGPPG (6/6)	CUTICLE COLLAGEN 40 (29.3 kDa)	CAENORHABDITIS	P34804†	AAA17726	7.4	21.3
7	HXAXXLGGG						
8	N-BLOCKED						
9	N-BLOCKED						
10	ATKKAVAVLKG (11/11)	SUPEROXIDE DISMUTASE, CHLOROPLAST PRECURSOR (21.3 kDa)	RICE	P93407†	BG046220	6.8	18.8
11	N-BLOCKED						
12	N-BLOCKED						
13	N-BLOCKED						
14	N-BLOCKED						
	GAPADPEDVR[b)]						
15	N-BLOCKED						
	ACYQSSLEQP	PUTATIVE CYTOCHROME (GIBBERELLIN) OXIDASE ASSEMBLY PROTEIN	RICE	AAK38309‡	AC079029	7.8	50.5
16	N-BLOCKED						
17	N-BLOCKED						
18	N-BLOCKED						
19	N-BLOCKED						
20	N-BLOCKED						
21	N-BLOCKED						
22	STAEMPAEPD	PUTATIVE PROTEIN	ARABIDOPSIS	A144482		4.2	24.2
23	APGVLYPQLA (10/10)	3-ISOPROPYLMALATE DEHYDROGENASE (38.9 kDa)	AZOTOBACTER	P96192†	CAA72146	7.8	28.4
24	N-BLOCKED						
25	EYGREDAXX						
26	N-BLOCKED						
27	N-BLOCKED						
	AKNYPVVSA	UNKNOWN PROTEIN FROM 2D-PAGE OF ETIOLATED COLEOPTILE (19.4 kDa)	MAIZE	P80627†	BE512153	5.9	26.5
28	VKAVVVLGSSEFPG	SUPEROXIDASE DISMUTASE (CU-ZN) 1 (15.1 kDa)	RICE	P28756†	BAA00799	7.9	21.6
29	N-BLOCKED						
	AGVVPKY						

Table 1. (Continued).

Spots no.	Sequence	Homologous protein	Species	Accession no.	EST Acc. no.[c]	pI	MW (kDa)
30	N-BLOCKED						
31	N-BLOCKED						
32	N-BLOCKED						
33	SVSXAQVEVSTWEX	CHLOROPLAST 30S RIBOSOMAL PROTEIN S7 (14.6 kDa)	RICE	P05424†	BAB47078	8.1	20.1
34	VGAGAVLASSEGVK	SUPEROXIDE DISMUTASE (CU-ZN) 2 (14.9 kda)	RICE	P28757†	BAA00800	7.9	18.9
35	N-BLOCKED						
36	PGDGAAAAAAAEPP	ELONGATION FACTOR 1-BETA (23.6 kDa)	RICE	P29545†	BAA02253	5.3	38.3
37	N-BLOCKED						
	AGVDD	DIHYDROFLAVONAL-4-REDUCTASE (38. 4 kDa)	BARLEY	P51106†	AAB20555	4.1	28.2
38	N-BLOCKED						
39	N-BLOCKED						
40	N-BLOCKED						
41	N-BLOCKED						
42	N-BLOCKED						
	AGVYLVA						
43	N-BLOCKED						
44	N-BLOCKED						
45	N-BLOCKED						
	DPYTVDYMHV	BETA-EXPANSIN EXPB2	RICE	AAK15440†	AC037426	8.1	42.1
46	N-BLOCKED						
	PEYPPDXD	HYPOTHETICAL PROTEIN	MAIZE	S05586‡		7.8	38.2
47	N-BLOCKED						
	PAIPAEE (7/7)	RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT CHLOROPLAST GENE ENCODING CHLOROPLAST PROTEIN	ONOCLEOPSIS	AAB41664‡	U62033	7.8	36.5
48	N-BLOCKED						
49	N-BLOCKED						
	STGVFTDKDK	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (36.5 kDa)	RICE	Q42977	AU062734	7.9	36.8
50	N-BLOCKED						

[a] *N*-terminal amino acid sequence
[b] Internal amino acid sequence.
[c] The EST numbers refer to the database entries with which the protein was identified (DDBJ or NCBI)
† Accession number is SWISS-PROT (http://www.expasy.ch/spot) database
‡ Accession number is NCBI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) database

compared with rice proteins may reflect partially their genomic and cDNA sequencing work in rice.

The simple new method developed in this study was effective for deblocking many proteins into amino acid, if sequenceable amounts is available. Since the proteins were used internal sequence analysis by Cleveland peptide mapping techniques (Hirano *et al.*, 1992; Komatsu *et al.*, 1993; Zhong *et al.*, 1997) and another (Tsugita *et al.*, 1994), this method is effective for the formyl group proteins. Cleveland gels followed by transfer to PVDF membranes and direct sequencing eliminate the need for manipulating small amounts of peptides in solution, allowing the sequencing of trace proteins that are amino-terminally modified. In addition, the cleavage with the *Staphylococcus aureus* V8 protease within the gels matrix have the greater specificity and efficiency for the rice embryo than proteolytic reactions in solution.

Peptide mapping has been widely used to show relationships between proteins (different proteins or the same protein from different sources) and to characterize structural changes in proteins (post-translational modifications such as glycosylation and phosphorylation) (Pandy & Mann, 2000). A Cleveland peptide map was generated for the rice young microspore stage and identified the amino-terminal amino acid sequence for the peptide. The peptide map, together with the entire primary sequence, which will soon become available through molecular biology and proteome techniques, is expected to be valuable for the characterization of protein functions and localization of modifications within the formyl group and the latter opens the pyroglutamyl group from the blocked.

It was found that green young microspore stage contained oxygen-evolving enhancer protein 2, chloroplast protein and mitochondrial protein. The presence of these photosynthetic enzymes in the green young microspore stage is normal as expected. However, as the electron transport chains of mitochondria and of chloroplasts are some of the well-known sources of active oxygen species (AOS) in plants, the cells contained antioxidant enzymes and metabolities to counteract these damaging AOS (Prasad, 1996).

The sequence comparisons showed sixteen proteins characterized in rice with 88~100% identity (Table 1). In rice, however, a great proportion of novel or functionally uncharacterized proteins expressed at the young microspore stage was found.

The usual tools were used to query the databases (FASTA and BLAST). For identified spots, an average match, over an average sequence length of 10 amino acids, was 85 %. The probability that these scores were obtained by chance is low, but two points are deserved for the further discussion. One-third of the identified spots, not all of the microsequences obtained, were found to be matched with the database entry. This is particularly true for the spots matching with partially sequenced ESTs. The incomplete matches might also mean that these spots correspond to isoforms not present in the databases.

In conclusion, the 2-DE based on proteome analysis has been successfully used to detect and characterize marker proteins that are idiotypic for a specific plant stress in physiological process or pathological process of a organelle or tissue. Recently, general proteomes have been constructed for numerous organelle and tissue with high success in protein identification. In this study, we have begun to establish the proteome of rice spikelets proteins. Of 250 spots detected in a young microspore stage fraction, sequences of 50 were obtained. In generating rice proteome research, it is necessary, however, to identify higher percentage (70%) of blocked *N*-terminal of rice embryo proteins. For this reason, it is necessary to develop novel techniques for high-throughput rice proteome research that allow for much increased starting amounts while permitting large-scale quantitative comparison of protein expression.

ACKNOWLEDGMENT

This work was supported by Chungbuk National University Grant in 2006.

REFERENCES

Cleveland, D. W., S. G. Fischer, M. W. Kirscher, and U. K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulphate and analysis by gel electrophoresis. J. Biol. Chem. 252 : 1102-1105.

Futcher, B., G. I. Latter, P. Monardo, C. S. McLaughlin, and J. I. Garrels. 1999. A sampling of the yeast proteome. Mol. Cell. Biol. 19 : 7357-7368.

Gygi, S. P., Y. Rochon, B. R. Franza, and R. Aebersold. 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19 : 1720-1730.

Hirano, H. 1982. Varietal differences of leaf protein profiles in mulberry. Phytochemistry 21 : 1531-1518.

Hirano, H., H. Takakura, F. Sakiyama, and S. Tsunasawa. 1992. Deblocking and subsequent microsequence analysis of Na-blocked proteins electroblotted onto PVDF membrne. J. Biochem. 111 : 754-757.

Hirano, H. and T. Watanabe. 1990. Microsequencing of proteins electrotransferred onto immobilizing matrices from polyacrylamide gel electrophoresis: Application to an insoluble protein. Electrophoresis 11 : 573-580.

Islam, N., S. Inanaga, N. Chishaki, and T. Horiguchi. 1996. Effect of N top dressing on the storage protein in Japonica and Indica rice. Cereal Chem. 73 : 571-573.

Komatsu, S., H. Kajiwara, and H. Hirano. 1993. A rice protein library: a data-file of rice proteins separated by two-dimen-

sional electrophoresis. Theor. Appl. Genet. 86 : 935-942.

O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250 : 4007-4021.

Pandy, A. and M. Mann. 2000. Proteomics to study genes and genomes. Nature 405 : 837-846.

Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann. 1996. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA. 93 : 14440-14445.

Prasad, T. K. 1996. Mechanisms of chilling-induced oxidative stress injury and tolerance in developing maize seedlings: changes in antioxidant system, oxidation of proteins and lipids, and protease activities. Plant J. 10 : 1017-1026.

Pearson, W. R. and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA. 85 : 2444-2448.

Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. miya, E. Murayama-Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala, and Y. Minobe. 1994. Toward cataloguing all rice cDNA from a callus cDNA library. Plant J. 6 : 615-624.

Tsugita, A., T. Kawakami, Y. Uchiyama, M. Kamo, N. Miyatake and Y. Nozu. 1994. Separation and characterization of rice proteins. Electrophoresis 15 : 708-720.

Varshavsky, A. 1996. The *N*-end rule' functions, mysteries, uses. Proc. Natl. Acad. Sci. USA 93 : 12142-12149.

Velculescu,V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. Science 270 : 484-487.

Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann. 1996. Femtomole sequencing of proteins from polyacrylamise gels by nano electrospray mass spectrometry. Nature 379 : 466-469.

Zhong, B., H. Karibe, S. Komatsu, H. Ichimura, Y. Nagamura, T. Sasaki, and H. Hirano. 1997. Screening of rice genes from a cDNA catalog based on the sequence data-file of proteins separated by two-dimensional electrophoresis. Breed. Sci. 47 : 245-251.