Available online at www.sciencedirect.com
SCIENCE DIRECT®
PHYTOCHEMISTRY
Phytochemistry 65 (2004) 1671–1681
www.elsevier.com/locate/phytochem

Review

Rice proteomics: recent developments and analysis of nuclear proteins

Md. Monowar Karim Khan, Setsuko Komatsu *

Department of Molecular Genetics, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba 305-8602, Japan

Received 25 December 2003; received in revised form 6 April 2004
Available online 6 May 2004

Abstract

Rice is the most important cereal crop in Asia, and is considered as a model cereal plant for genetic and molecular studies. An immense progress has been made in rice genome sequence analysis during the last decade. This prompted the researcher to identify the functions, modifications, and regulations of every encoded protein. Proteome analysis provides information to predict the translation and relative concentration of gene products, including the extent of modification, none of which can be accurately predicted from the nucleic acid sequence alone. During the last couple of years, considerable researches were conducted to analyze rice proteome, and only recently a remarkable progress has been made to systematically analyze and characterize the functional role of various tissues and organelles in rice. In this review, the rice proteomic research has been compiled and also presented a comprehensive analysis of rice nuclear proteins. In rice nucleus, 549 proteins were resolved using 2D-PAGE. Among them, 257 proteins were systematically analyzed by Edman sequencing and mass spectrometry and identified 190 proteins following database searching (http://gene64.dna.affrc.go.jp/RPD/main.html). The identified proteins were sorted into different functional categories. In these data, the proteins involved in signaling and gene regulations dominated others, reflecting the role of nucleus in gene expression and regulation.

Keywords: *Oryza sativa* L.; Rice; Proteome; Nucleus; 2D-PAGE

Contents

* Corresponding author. Tel.: +81-29-838-7446; fax: +81-29-838-7464/7404.
E-mail address: skomatsu@affrc.go.jp (S. Komatsu).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.04.012

1. Introduction

Rice (*Oryza sativa* L.) is an ideal model plant among the monocot cereal crop species for genetic and molecular studies because of its smaller genome than those of other cereals (Devos and Gale, 2000). It is the main staple food for more than half of the world's population (Sasaki and Burr, 2000). To feed the increasing population of the world, rice draws the crucial attention to improve its yield and quality. The publications of draft genome sequences for *Oryza sativa* L. ssp. indica (Yu et al., 2002) and *Oryza sativa* L. ssp. japonica (Goff et al., 2002) and complete map-based sequence of chromosome 1 (Sasaki et al., 2002) and chromosome 4 (Feng et al., 2002) for *Oryza sativa* L. cv. Nipponbare, provide a rich resource for understanding the biological process in rice. The challenge ahead for the plant research community is to identify the functions, modification, and regulation of every encoded protein. The understanding of the biological functions of the novel genes is a more difficult proposition than obtaining just the sequences. This is because of the fact that the existing amount of information on amino acid sequences of known proteins in the database does not match the wealth of information on nucleotide sequences being generated through genome projects (Lockhart and Winzeler, 2000; Pandey and Mann, 2000). The analysis of proteins is the most direct approach to define the gene function. Proteome analysis provides information on functional genomics, including identification of open reading frames (ORFs) from genome sequences, determining the proteins for subcellular compartment localization, and identifying novel components involved in the components' biogenesis. Currently, proteomics is becoming an increasingly powerful tool for the investigation of complex cellular process and turning out to be a major international subject of research.

Over the past few years, significant progress has been made towards identifying and cataloguing of the proteins from rice tissues and organelles. Proteomics of rice embryo and endosperm (Komatsu et al., 1993); root (Zhong et al., 1997); green and etiolated shoot (Komatsu et al., 1999a); cultured suspension cells (Komatsu et al., 1999b); anther (Imin et al., 2001); leaf sheath (Shen et al., 2002), and different organelles such as Golgi (Mikami et al., 2001); mitochondria (Heazlewood et al., 2003) have opened an avenue to critically understand the functions of rice proteins. Tsugita et al. (1994) analyzed and identified 4892 proteins from nine tissues and one organelle of rice (leaf, stem, root, germ, dark germinated seedling, seed, bran, chaff, callus and chloroplast). A more detail proteome analysis of rice (*Oryza sativa* L. cv. Nipponbare) leaf, root, and seed has been reported (Koller et al., 2002). The recently constructed of rice proteome database website (http://gene64.dna.affrc.go.jp/RPD/main.html), provides an enormous information on the progress of rice proteome research (Komatsu et al., 2004). Proteome analysis of various tissues and organelles has shown diverse functional categories. Although many common proteins were identified and shared similar functions in different tissues and organelles, but majority of the proteins were tissue and organelle specific. These results encouraged the researchers to analyze the proteins from different tissues and organelles separately.

To better understand the proteome of rice, it is urgently needed to analyze the proteins of various tissues and organelles separately; those have not yet been analyzed. To date, no attempts have been made to analyze the proteins of rice nucleus. The nucleus is one of the most important organelles in living organism that regulates various biological activities, including gene expression. It stores genes on chromosomes, produces messages that code for protein, transport regulatory factors and also functions to many signaling responses. Unlike rice, few reports on the nucleus of *Arabidopsis* have been published and highlighted their molecular functions (Bae et al., 2003; Folta and Kaufman, 2000). Prediction of organelle specific proteins by different programs such as MitoProtII (Claros and Vincens, 1996), TargetP (Emanuelsson et al., 2000), and Predotar (http://www.inra.fr/Internet/Produits/Predotar/) are not yet completely possible due the lack of sufficient proteome information on various organelles. These programs predict thousands of nuclear-encoded plant proteins to be mitochondrial (Heazlewood et al., 2003). Gomez et al. (2003) used these software packages to predict organelle targeting and transit peptide proteolysis site and conclude that some programs (ChloroP and TargetP) reliably predict trafficking to the chloroplast, but

they variably predict the processing sites of the transit peptides and suggested an improvement in the datasets used to train the algorithms. The complete analysis of nuclear protein could clarify the veracity of such predictions. In this review, attempts were made to compile almost all the recent works on rice proteome to get a clear understanding on the progress in this area of research to date. Further, studies were made to systematically analyze the rice nuclear proteome.

2. Proteome analysis of rice tissues

Considerable efforts have been made to analyze the proteins from different tissues and organelles in rice (Table 1). 'A rice proteome library' constructed by Komatsu et al. (1993) is the pioneer report on the proteome analysis of rice tissues. Proteins from seed embryo, endosperm and leaf were isolated, and separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) (O'Farrell, 1975). More than 600, 100 and 150 major proteins were detected in embryo, endosperm and leaf, respectively by Coomassie brilliant blue (CBB) staining. Seven hundred proteins in leaf sheath were detected by silver staining. Eighty-five protein spots were selected from embryos, endosperms and leave and analyzed by a gas-phase protein sequencer. Among these 85 protein spots, 27 proteins were identified by N-terminal sequence analysis. The rest 70% proteins were considered to have a blocking group at the N-terminus. Then by in situ treatment with pyroglutamyl peptidase deblocked the N-terminus and identified 11 proteins. Furthermore, 23 proteins were analyzed for their internal sequences following Cleveland peptide mapping (Cleveland et al., 1977). Finally, the authors constructed data-files of rice embryo, endosperm and leaf and provided molecular weight and isoelectric point (p*I*) for each protein.

Tsugita et al. (1994) analyzed the proteins from nine different tissues of rice (leaf, stem, root, germ, dark germinated seedling, seed, bran, chaff and callus) and one organelle (chloroplast). From the above nine tissues, a total of 4616 protein spots were identified on 2D-gels, of which 118 proteins were analyzed, and among them 43 proteins could be sequenced. But due to lack of information on the rice genome and protein database at that time, only 16, 6, 3, 1 and 1 protein could be identified in leaf, callus, germ, stem and root, respectively. *N*-terminal sequencing could characterize only about 3% of the spots. The same group carried out another comparative study on the above nine tissues of rice and five tissues (seed, callus, stem, root and leaf) of *Arabidopsis* (Tsugita et al., 1996). About 5000 protein spots were separated by 2D-PAGE from both the nine tissues of rice and five tissues of *Arabidopsis*. More that 100 protein spots were analyzed for *N*-terminal sequences and 62 proteins in *Arabidopsis* and 51 new proteins in rice could be sequenced. Among the identified proteins, 40% were found to have common function between these two species. Their results indicated the presence of many common proteins among different species.

Zhong et al. (1997) analyzed the proteins from rice root and detected 292 protein spots by CBB staining. Seventy six proteins were analyzed following N-terminal and internal amino acid sequences. The amino acid sequences of these proteins were compared with those of PIR database (Pearson and Lipman, 1988) and a total of 42 proteins were identified by homology search in the Rice Genome Research Project cDNA catalog. The authors also described a simple approach to use database interrogation and homology search to find the putative cDNA corresponding to the protein that was sequenced. During the same year Hirano (1997) investigated proteins from endosperm and embryos of seed, leaves and roots following 2D-PAGE technique. A total of 278 proteins were resolved and N-terminal and internal amino acid sequences of 121 proteins were sequenced in these four tissues. While, sequences of only 51 proteins were found identical to those proteins reported in the database. This finding provided significant knowledge on the differential expression of proteins in different tissues.

Komatsu et al. (1999a) carried out a comparative proteome investigation on green and etiolated shoots of rice seedlings, and analyzed 85 proteins by a gas-phase proteins sequencer, out of 300 major protein spots detected on a 2D-gel by CBB staining. Among them, 21 and 39 proteins were identified by *N*-terminal and internal amino acid sequences, respectively. In this manner, the major proteins were identified to be involved in growth regulation in green and etiolated rice shoots of seedlings, although nine proteins were found to be common in both the tissues. During this year, this group of scientists conducted another proteome analysis of rice suspension cultured cells (Komatsu et al., 1999b). Onehundred three major protein spots from this cells were analyzed and determined the amino acid sequences of 20 and 32 proteins by N-terminal and internal amino acid sequence analysis, respectively. The identified proteins were searched for their functional categories and found to be involved in growth and development of rice cell suspension cultures. This work provided a basic idea on the major growth regulating proteins present in these cells.

Imin et al. (2001) used rice anthers at the young microspore stage for proteome analysis. This organ of plant is critical for plant production and is sensitive to various environmental stresses. Over 4000 proteins spots were detected on 2D-PAGE within the p*I* range 4–11 and molecular weight 6-122 kDa by silver staining. Among these 4000 proteins spots, they analyzed 273 proteins by N-terminal sequencing and mass spectrom-

Table 1
Proteome analysis of various tissues and organelles in rice

Tissue/organelle	Spots[a]	ID by Edman sequencing[b]	ID by mass tag	Experimental condition	Ref.
Tissue					
Embryo	600^{CBB}, 700^{Ag}				Komatsu et al. (1993)
Endosperm	100^{CBB}	27^{N}, 11^{U}, 23^{I}		2D-PAGE	
Leaf	150^{CBB}				
Leaf	4616^{Ag}	43^{Tot} (16^{M})		2D-PAGE	Tsugita et al. (1994)
Stem		(1^{M})			
Root		(1^{M})			
Germ		(3^{M})			
Dark					
Germinated					
Seedling					
Seed					
Bran					
Chaff					
Callus		(6^{M})			
Leaf	$\sim5000^{Ag}$	51^{N}		2D-PAGE	Tsugita et al. (1996)
Stem					
Root					
Germ					
Dark					
Germinated					
Seedling					
Seed					
Bran					
Chaff					
Callus					
Root	292^{CBB}	76^{N+I}, (42^{M})		2D-PAGE	Zhong et al. (1997)
Endosperm		121^{N+I}, (51^{M})		2D-PAGE	Hirano (1997)
Embryos					
Leaf					
Root					
Green and etiolated shoot	$\sim300^{CBB}$	21^{N}, 39^{I}		2D-PAGE	Komatsu et al. (1999a)
Cultured suspension cells	$\sim300^{CBB}$	20^{N}, 32^{I}		2D-PAGE	Komatsu et al. (1999b)
Anther	$\sim4000^{Ag}$	273^{N+MS}(62^{M})		2D-PAGE	Imin et al. (2001)
Leaf	$\sim2600^{Ag}$			2D-PAGE after PEG fractionation	Kim et al., 2001
Leaf sheath	352^{CBB}	44^{N+I} (31^{M})	59 (59^{M})	2D-PAGE	Shen et al. (2002)
Embryo	700^{CBB}	31^{N}, 69^{I} (28^{M})	150 (46^{M})		Woo et al. (2002)
Leaf			2528^{Tot}	2-DE and	Koller et al. (2002)
Root			(2251^{M})	MudPIT	
Seed					
Leaf sheath	431^{CBB}	7^{N}, 6^{I}	66^{M}	2D-PAGE	Tanaka et al. (2004a,b)
Root	508^{CBB}	38^{N}	35^{M}	Gibberellin	
Cultured suspension cells	962^{CBB}	50^{N}	90^{M}	Regulation	
Organelle					
Chloroplast	276^{Ag}	13^{N} (7^{M})		2D-PAGE	Tsugita et al. (1994)
Golgi					Mikami et al. (2001)
Mitochondria		$232^{(MS)}$	$91^{(M)}$	BN-PAGE	Heazlewood et al. (2003)
		$170^{(LC/MS)}$	$45^{(M)}$	LC–MS	
Nucleus	549^{CBB}	17^{N}, 22^{I} (30^{M})	178 (160^{M})	2D-PAGE	This study

U – unblocked peptide sequence; I – internal peptide sequence; M – identification made by homology search with peptide sequence.
[a] CBB – Coomassie brilliant blue stain, Ag – sliver stain.
[b] N – N-terminal peptide sequence.

etry. And through database matching and PMF analysis 62 different proteins were identified. The major proteins identified consisted of house keeping proteins, heat shock proteins, many putative proteins, including a 20S proteasome, Jun activated domain binding proteins and a newly identified translationally controlled tumor protein. They have constructed a proteome reference map for this organ and have made available at http://semele.anu.edu.au/2d/2d.html. This rice anther proteome map presents an insight into the normal microspore development. Moreover, information provided in this database also serves as a good reference for similar studies in other plants.

An improved fractionation method proposed by Kim et al. (2001) for the analysis of rice leaf proteins put forward one step in this area of research. The high abundance of RuBisCO in rice leaf often limits the analysis of other low abundant proteins. They have resolved this drawback by the development of a simple extraction and fractionation technique with polyethylene glycol (PEG) to analyze rice leaf proteins. This fractionation technique helped them to analyze 2.7 times more well separated protein spots compared with single-step 2-DE analysis. Following this improved technique, at least 2600 well-separated protein spots from leaves could be analyzed. The PEG fractionation technique, thus found to be superior to the typical used ammonium sulfate or sucrose gradient fractions. Technical modification on the existing technology made by this group contributes significantly for proteome analysis of complex tissues.

In 2002, an efficient proteome analysis of leaf sheath from rice seedling was performed using 2D-PAGE followed by Edman sequencing and matrix-assisted laser desorption ionizaton time-of-flight (MALDI-TOF) mass spectrometric analysis (Shen et al., 2002). Three hundred fifty two protein spots were detected by image analysis in this tissue and the amino acid sequences of 44 out of 84 major selected proteins were determined; and a clear function of 31 proteins were assigned but for 12 proteins, no function could be assigned. Further, by using MALDI-TOF-MS they were able to identify 59 proteins having similarity with the database. The physiological significance of the proteins identified from the N-terminal and internal amino acid sequences were determined. This work demonstrates that a combination of 2D-PAGE, Edman sequencing, MALDI-TOF-MS and MS/MS is very important for the efficient analysis of rice leaf sheath proteins. Almost in parallel, an efficiently peptide mapping technique has been reported and illustrated its application to identify embryo proteins for rice proteome analysis (Woo et al., 2002). Seven hundred proteins were separated on 2D-gels and detected by CBB staining. One hundred embryo proteins were selected for analysis, of which the N-terminal amino acid sequences of 31 proteins could be done. Rest 69 proteins were *N*-terminally blocked and analyzed for their internal sequences following Cleveland peptide mapping method. Out of 100 analyzed proteins, only 28 proteins showed sequence similarity to the proteins with known function in the database. They also analyzed another 150 proteins by peptide mass fingerprinting using MALDI-TOF-MS and could identify only 46 proteins. Rest of the proteins could not be identified due to the lack of complete nucleotide sequence of rice genome and the information provide in the protein database to date in not enough to identify all the proteins. Nevertheless, their work provides an improved sample preparation for peptide mapping after 2D-PAGE, which is very useful for rice proteome analysis.

Koller et al. (2002) have presented a detailed and systematic proteome work on rice tissues. The proteome from leaf, root and seed from rice were extensively analyzed using 2D-PAGE and multi-dimensional protein identification technology (MudPIT). Following these two techniques, they could detect and identify a total of 2528 unique proteins (6296 peptides) in these tissues. Of these 2528 proteins, 2251 proteins were identified with peptides that uniquely identify the proteins. The other 277 proteins could not be identified, because the peptides that matched to these proteins also matched to other proteins. The identified proteins were sorted into 16 different groups according to their functional categories. The most abundant category was reported to be occupied by proteins with as yet unidentified function and proteins with very low or no detectable homology to other predicted proteins in the database (32.8%). The second most abundant class of proteins was classified as being involved in metabolic processes (20.8%), followed by protein destination (8%), cell rescue, defense, cell death and ageing (6%), protein synthesis (5%), energy (5%), cellular organization (5%), retrotransposons and plasmid proteins (5%), transport facilities (1%), development (1%), and intracellular transport (1%) etc. It is important to note that majority of the proteins identified by this group showed a tissue-specific expression pattern; unlike 189 proteins (7.5%) those were expressed in all the 3 tissues analyzed. A tissue-specific expression of metabolic pathways has also been reported here using proteomic approach. These results indicate that the combination of multidimensional proteomic technologies with quantitative methods allow the integration of mRNA and protein expression data, and endorse to understand the complex metabolic networks in plants.

Recently, Tanaka et al. (2004a) have conducted an well-organized proteome analysis of rice leaf sheath, root and cultured suspension cells using 2D-PAGE, Edman sequencing, mass spectrometry followed by database searching. They also investigated the gibberellin (GA) regulated proteins in these tissues. The identified proteins were 79 of the 431 spots detected on 2D-gel by CBB staining in leaf sheath, 73 of the 508 spots in root,

and 140 of the 962 spots in cultured suspension cells. Protein data-files were constructed and divided the proteins according to their functional categories. The GA regulated proteins were also determined from the data-files and reported that 8, 20, and 15 proteins were regulated by GA in leaf sheath, root and cultured suspension cells, respectively. The rice tissue proteome data-files constructed by this group were found to be very useful for rapid assessment of any changed proteins by GA. Komatsu et al. (2004) has constructed a rice proteome database web-site (http://gene64.dna.affrc.go.jp/RPD/main.html) that provides enormous information on proteome analysis of 15 different tissues and organelles of rice. This is a milestone for functional analysis of rice genome and might serve as a useful site for the researchers working on proteome.

3. Proteome analysis of rice organelles

Rice subcellular compartment proteomics provide available information on functional genomics, which include the determination of the open reading frames from genome sequences, proteins for sub cellular compartment function, and identify novel components involved in the compartments biogenesis. To date, a small number proteome research has been carried out on various organelles in rice (Table 1).

Analysis of proteins from rice chloroplast is the first organelle proteome work (Tsugita et al., 1994). A total of 276 protein spots were resolved on 2D-gel, of which only 16 proteins were analyzed for their *N*-terminal sequences and 13 proteins could be sequenced. Out of the sequenced proteins, seven proteins could be identified and six proteins remained unknown. This is a preliminary but the first report on rice chloroplast proteome analysis. Their report provides a brief idea on the proteins in rice chloroplast and encouraged the researchers to study this organelle in detail for better understanding the critical role of chloroplast in rice plant.

Mikami et al. (2001) separated the Golgi complex from rice. Golgi complex is a very important multifunctional organelle in rice where many important biochemical process, particularly biosynthesis of complex cell surface polysaccharides and the processing and modification of glycoproteins take place. The structure and functional unit of the plant Golgi complex is thought to be the Golgi stack, the *trans*-Golgi network and the Golgi matrix. The distinct compartments of the Golgi from rice and tobacco were separated following sucrose density gradient centrifugation technique with EDTA or $MgCl_2$. At least 4 distinct compartments in the rice Golgi complex have been reported. Rice microsomal proteins were also isolated following SDS–PAGE technique. Using lectin and some specific antibody, they determined the sedimentation behaviors of compartments specific proteins in rice Golgi complex. These results facilitate to isolate different components from rice Golgi complex and further proteome analysis. Furthermore, an improvement on 2-DE technology for the separation of rice organelle membrane proteins has been reported (Mikami et al., 2002). The cytosolic and membrane associated proteins from rice cells were separated and compared by two different 2D-PAGE methods, namely isoelectric focusing (IEF)/SDS–PAGE and non-equilibrium pH gradient electrophoresis (NEPHGE/SDS–PAGE). It has been suggested that NEPHGE/SDS–PAGE is very useful for the proteome analysis of rice membrane associated proteins, but IEF/SDS–PAGE of the cytosolic fraction showed more resolution.

Recently, a comprehensive report on rice mitochondrial proteome has been presented (Heazlewood et al., 2003). The separated the purified mitochondrial proteins by IEF/SDS–PAGE, blue-native PAGE, and reverse phase high performance liquid chromatography (LC). Following these separation techniques, they could detect highly basic and acidic proteins (isoelectric point 4.0–12.5), and proteins over a large molecular mass (6.6–252 kDa). About 250 mitochondrial protein spots were resolved by IEF/SDS–PAGE technique and 145 spots were analyzed by MS/MS. They could identify 80 proteins as the products of 63 predicted gene sequence from rice. While, the membrane fraction of rice mitochondria that contains a large number of hydrophobic proteins was separated by blue-native SDS–PAGE and detected 100 protein spots. Out of them, 89 protein spots were analyzed and 57 proteins were identified as the products of 49 rice genes. Sequence comparison allowed them to assign functions to 85 proteins. Further, complete digestion of mitochondrial proteins with trypsin yielded a peptide mixture that was analyzed directly by reverse-phase LC. This provided 170 MS/MS spectra that matched 72 sequence entries from open reading frame and expressed sequence tag database. The authors reported the first direct identification of pentatricpeptide repeat protein in plant mitochondrial proteome analysis. Their works provide a comprehensive mitochondrial proteome analysis including the tools required for extensive plant proteome analysis.

4. Proteome analysis of rice nucleus

4.1. Results

4.1.1. Isolation of purified nuclei from rice cultured suspension cells

The integrity of a subcellular proteome, such as nuclei, is largely dependent on the purification of the isolated compartment away from other cellular contaminants. The separation of high-purity nuclei from plant is a

difficult task. A series of fractionation process were followed to obtain purified nuclei. Initially, 2.0 M sucrose density gradient (Morre and Andersson, 1994) and percoll density gradient (Folta and Kaufman, 2000) centrifugation techniques were adopted to isolate nuclei from rice cultured suspension cells. The modified sucrose gradient method (described in this study) was found more rapid and efficient for the isolation of nuclei from rice cultured suspension cells. Photomicrograph of nuclei-enriched fraction is presented in Fig. 1(a). The nuclei were uniform spheres with an average diameter of approximately 20 μm. The nuclear proteins were prepared from the purified nuclei using lysis buffer (O'Farrell, 1975). The protein profile of the nuclear fraction on SDS–PAGE appeared distinct from that of the other fractions (Fig. 1(b)). The purity of the isolated nuclear fraction was evaluated by Western blot analysis using Histone H1, a specific antibody for nuclear protein. Histone H1 was found in the nuclear fraction, but not in the supernatant fraction (Fig. 1(b)), suggesting that the preparation is enriched in nuclear proteins.

4.1.2. Identification of nuclear proteins

Isolated nuclear proteins were separated by 2D-PAGE using IEF and IPG tube gels (Hirano et al., 2000). The samples were analyzed in more than five times. Computer analysis of the Image-Master 2D Elite software of the CBB stained gels reproducibly revealed 549 different spots in the pH range 4–10. After 2D-PAGE, proteins were analyzed by Edman sequencing and mass spectrometry. Among the 549 proteins detected in the nucleus of rice, 257 protein spots those were more abundant in the 2D-gels were selected for analysis. Initially, the proteins separated by 2D-PAGE were electroblotted onto a polyvinylidene difluoride (PVDF) membrane and 100 strongly visible protein spots were selected for analysis by Edman sequencing. The N-terminal amino acid sequences of 17 out of 100 proteins were determined in this manner and 83% of the proteins were N-terminally blocked. The internal amino acid sequences of 22 proteins were determined by sequence analysis of peptides obtained by peptide mapping (Cleveland et al., 1977) with *Staphylococcus aureus* V8 protease. Among the amino acid sequences of 39 proteins determined from N-terminal and internal amino acid sequence analysis, the sequences of nine proteins did not match to the sequences of known proteins/genes in the database. Thus, 30 nuclear proteins could be identified by Edman sequence analysis. Then 178 proteins were systematically analyzed by MALDI-TOF-MASS and a total of 160 proteins could be identified, other 18 proteins did not have any homologies to any rice proteins/genes in the available database. The detailed list of the identified nuclear proteins is provided in Rice Proteome Database web-site, which is available at http://gene64.dna.affrc.go.jp/RPD/main.html (see Fig. 2).

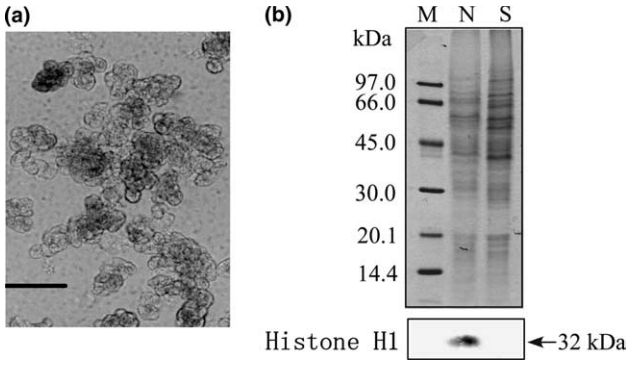

Fig. 1. Isolation of rice nuclear proteins: (a) micrograph of isolated rice nuclei. Bar, 50 μm; (b) Western blot analysis of fractionated proteins with anti-Histone-H1 antibody. M, low molecular marker; N, nuclear fraction from sucrose density gradient centrifugation; S, supernatant fraction from sucrose density gradient centrifugation.

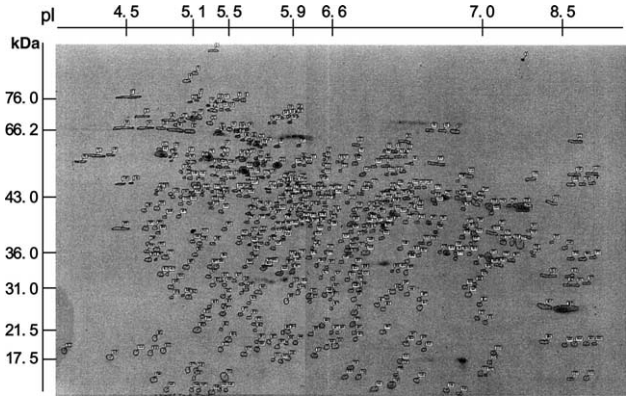

Fig. 2. Two-dimensional polyacrylamide gel electrophoresis map of rice nuclear proteins. Proteins were extracted from the purified nuclei of rice cultured suspension cells and separated by 2D-PAGE with IEF and IPG in the first dimension and SDS–PAGE in the second dimension. The 2D maps separately constructed from IEF and IPG tube gels were overlapped at around p*I* 6.0. After detection by CBB staining, protein spots were analyzed by Image-Master 2D Elite. Numbers indicate the different nuclear protein spots and the spots, which are not numbered, are the molecular markers. The p*I* and relative molecular weight of each protein were determined using 2D-PAGE marker.

4.1.3. Functional classification of nuclear proteins

To understand the function of the nuclear proteins, the identified proteins were sorted into different categories as shown in Fig. 3. Of the 190 proteins identified following Edman sequencing and mass spectrometric analysis, the most abundant category was classified as being involved in signal transduction and gene regulation (29%). Proteins with no assigned function occupied the second most abundant class (26%), as the function of these proteins are not yet identified or clearly known and are therefore considered to be rice nuclear-specific

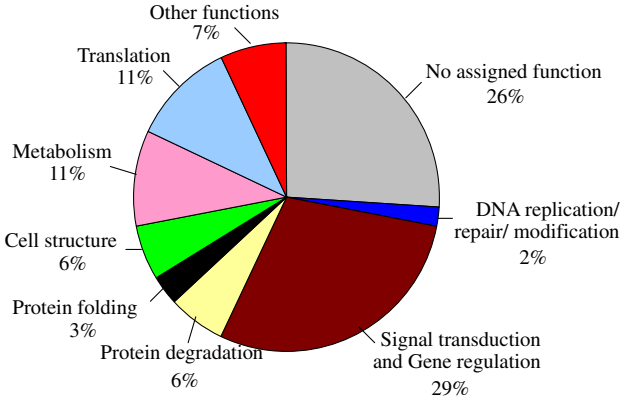

Fig. 3. Assignment of the identified nuclear proteins to functional categories. The percentages of proteins in each functional category from the total number (190) of identified proteins are shown.

proteins. A high percentage (10%) of proteins were identified to be involved in metabolism, which include RNA and carbon metabolism. A significant fraction of the proteins were involved in housekeeping functions, such as cell structure (6%), translation (11%), and DNA replication/repair/modification (2%). The identified nuclear proteins were also found to be involved in protein degradation (6%) and protein folding (3%). Proteins categorized in other functions, include the proteins involved in defense, energy production and cell growth etc. Thus, using 2D-PAGE, Edman sequencing, and mass spectrometry, followed by database searching, nuclear proteins were comprehensively identified, and observed that the proteins were involved in a variety of functions.

4.2. Discussion

The gene sequence of rice is in immense progress (Sasaki et al., 2002) and complete rice genome for different cultivars is going to be published soon. Meanwhile, proteome projects were initiated to systematically characterize gene products of defined organelles from rice (Heazlewood et al., 2003; Tanaka et al., 2004b). The present paper reports the initiation of rice proteome project. An important criterion for a specific organelle proteome is the purity of the component to be analyzed. The rice nucleus is one of the most complex and very important organelles in rice cell. It consists of multiple compartments including the nucleus and carries out a variety of processes fundamental to cell function (Lamond and Earnshaw, 1998). The rice nuclei isolated in this study was checked under the light microscope as well as Western blot analysis. The nuclei were uniform spheres with an average diameter of approximately 20 μm. Pea nuclei were also observed uniform spheres with an average diameter of about 20–30 μm, but the *Arabidopsis* nuclei were smaller in diameter (5–0 μm) and non-spherical (Folta and Kaufman, 2000). Data suggests that the isolated nuclei were highly purified. Hence, proteins from the purified nuclei were extracted and used for further proteome analysis.

The combination of 2D-PAGE and MS is a better approach to study complex patterns of gene expression at the proteome level (Pandey and Mann, 2000), and provides both higher sensitivity and higher throughput than is possible with Edman sequencing. A total of 549 rice nuclear proteins were resolved within the p*I* value 4.0–10. This is in consistence with the *Arabibopsis* nuclear proteins, reported by Bae et al. (2003), where they detected around 544 protein spots in the p*I* range of 3–10. Among the proteins detected, 257 proteins were analyzed using Edman sequencing and MS analysis and identified 190 proteins following database searching. This is the first comprehensive report on rice nuclear proteome to our knowledge. The identified proteins were classified according to their functions (Fig. 3). The role of rice nucleus can be plainly understood from this graph. The proteins were found to be involved in different cellular functions, e.g., signaling, gene regulation, structure, metabolism, translation, protein degradation and folding. It is interesting to note that proteins involved in signaling and gene regulation dominated others, reflecting the role of nucleus in gene expression and regulation. While, in other organelles, like chloroplast and mitochondria, the largest percentage of proteins were reported to be involved in energy production, either in electron transport or in ATP production (Millar et al., 2001; Peltier et al., 2000).

4.3. Experimental

4.3.1. Extraction and separation of rice nuclear proteins

Nuclei were prepared from rice cultured suspension cells following the sucrose density gradient method described by Morre and Andersson (1994) with some modification. Rice cultured suspension cells used for the isolation of nuclei was cultured in N6 liquid medium (Murashige and Skoog, 1962) supplemented with 1 ml l^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D) under continuous shaking in an incubator at 22 °C. All the steps for the isolation of nuclei were performed on ice or at 4 °C. Approximately 3 g of rice cultured suspension cells were taken in a glass mortar and five ml homogenization medium containing 50 mM HEPES (pH 7.4), 10 mM KCl. 1mM EDTA, 10 mM ascorbate, 0.1% bovine serum albumin, 20 mM dithiothreitol, and 400 mM sucrose was added and homogenized by a pestle. The homogenates were passed two times through a double layer Miracloth (Calbiochem, Darmstadt, Germany). The homogenate was transferred to a 15 ml falcon tube (Becton Dickinson Labware, Franklin Lakes, NJ, USA)

and centrifuged at 1000 g for 10 min in an AG-6512C rotor. The pellet was resuspended in the homogenization medium and layered on top of a 2.0 M sucrose cushion prepared in 37.5 mM Tris–maleate (pH 6.5), 5 mM $MgCl_2$, and 1% dextrin T500 (Pharmacia, Uppsala, Sweden) and centrifuged at 50,000*g* for 30 min. The supernatant solution was removed carefully, and the pellet was resuspended again in the above homogenization medium. The suspension was further layered on top of the 2.0 M sucrose cushion and centrifuged at 50,000*g* for 30 min. The pellet containing the nuclei was observed under a microscope (Nikon microphot-FXA, Yokohama, Japan). The nuclei enriched pellet was employed for the isolation of nuclear proteins using lysis buffer (O'Farrell, 1975). Purity of the nuclear proteins was tested following Western blot analysis using anti-Histone H1 antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA).

4.3.2. Two-dimensional polyacrylamide gel electrophoresis

The prepared nuclear proteins were separated in the first dimension by isoelectric focusing (IEF) or immobilized pH gradient (IPG) tube gel (pH 6–10, Daiichi-Kagaku, Tokyo, Japan) and in the second dimension by SDS–PAGE. IEF gel tube solution consisted of 8 M urea, 3.5% acrylamide, 2% NP-40, and an equal mixture of 2% Ampholines (pH 3.5–10 and 5–8). Electrophoresis was carried out at 200 V for 30 min, followed by 400 V for 16 h and 600 V for 1 h. For IPG gel electrophoresis, samples were applied to acidic side of gels and electrophoresis was carried out at 400 V for 1 h, followed by 1000 V for 16 h and 2000 V for 1 h. SDS–PAGE in the second dimension was performed using 15% polyacrylamide gel and 5% staking gel at a constant current of 35 mA. The gels were stained with CBB and p*I* and molecular mass were determined by the 2D-PAGE Marker (Bio-Rad, Richmond, CA, USA). After staining, gels were scanned using a flatbed scanner, and the data were analyzed using Image-Master 2D Elite software (Amersham Biosciences, Piscataway, NJ, USA).

4.3.3. N-terminal and internal amino acid sequence analysis

The proteins separated by 2D-PAGE gel were electroblotted onto a PVDF membrane (Pall Bio Support Division, Port Washington, NY, USA) using a semi-dry transfer blotter (Nippon Eido, Tokyo, Japan), and visualized by staining with CBB. The protein spots were excised from the PVDF membrane and directly subjected to Edman degradation on a gas-phase protein sequencer (Procise 494, Applied Biosystems, Foster City, CA, USA). The N-terminally blocked proteins were further analyzed following Cleveland peptide mapping technique (Cleveland et al., 1977). Following separation by 2D-PAGE, gel pieces containing protein spots were excised out and the proteins were electroeluted from the gel pieces using an electrophoretic concentrator (Nippon Eido) at 2 W constant power for 2 h. After electroelution, the protein solution was dialyzed against deionized water for 2 days and lyophilized. The protein was dissolved in 20 μl of an SDS sample buffer containing 0.5 M Tris–HCl (pH 6.8), 10% glycerol, 2.5% SDS, and 5% 2-mercaptoethanol and applied to a sample well in an SDS–PAGE gel. The sample solution was overlaid with 20 μl of a solution containing 10 μl of *S. aureus* V8 protease (0.1 $\mu g\,\mu l^{-1}$) (Pierce, Rockford, IL, USA) and 10 μl of the SDS sample buffer. Electrophoresis was carried out until the sample and protease were stacked in the staking gel, interrupted for 30 min to digest the protein. Electrophoresis was further continued and the digests were electroblotted onto a PVDF membrane and subject to amino acid sequencing as above. The amino acid sequences were compared with protein sequences in a SWISS-Prot, PIR, Genpept and PDP database using the FastA sequence alignment program.

4.3.4. Matrix-assisted laser desorption ionization time-of-flight mass spectrometric analysis

The CBB stained protein spots were excised from gels, washed with 25% methanol and 7% acetic acid for 12 h at room temperature, and destained with 50 mM NH_4HCO_3 in 50% methanol at 40 °C for 1 h. After drying in vacuum, the gel spots were reduced with 10 mM EDTA, 10 mM dithiothreitol, and 100 mM NH_4HCO_3 at 60 °C for 1 h. The gels were dried in vacuum and incubated in an alkyline solution containing 10 mM EDTA, 40 mM iodoacetamide, and 100 mM NH_4HCO_3 at room temperature for 30 min in the dark. The gel pieces were minced, dried in vacuum and digested in 10 mM Tris–HCl (pH 8.0) containing 1 pM trypsin (Sigma, St. Louis, MO, USA) at 37 °C for 12 h. The digested peptides were extracted from the gel slices with 0.1% trifluoroacetic acid in 50% acetonitrile/water for three times. The peptide solution, thus obtained, was dried up and reconstituted with 30 μl of 0.1% TFA in 5% acetonitrile/water, and then desalted by ZipTipC18^{TM} pipette tips (Millipore, Bedford, MA, USA). The purified peptide solutions (2 μl) were mixed with equal volume of matrix solution (10 $mg\,ml^{-1}$ α-cyano-4-hydroxycinnamic acid, 0.3% trifluoroacetic acid and 50% acetonitrile) and air-dried on a plate for analysis using MALDI-TOF-MASS (Voyager-DE PR, Applied Biosystems, Framingham, MA, USA). Calibration was carried out using a standard peptide mixture. The mass spectra were subjected to sequence database search with Mascot software (Matrix Science Ltd., London, UK).

5. Conclusion

Rice proteome analysis began its way in 1995 with the separation and cataloging of proteins from different tissues. The remarkable progress achieved in protein identification and functional analysis during the last 5 years, makes objectives beyond the simple cataloguing of proteins a realistic aim. This progress was made possible due to some major technical advances of proteome analysis including the progress in rice genome sequence analysis. The various rice proteome work outlined here, reflects that proteome is growing in scope and importance in the field of rice functional genomics. The comprehensive proteome analysis of rice nucleus presented here provides an insight into the functional role of the organelle. Still there are a number of rice tissues and organelles that need to be addressed comprehensively to improve and develop rice proteomics to its full. Besides the thousands of proteins that remain to be discovered following the sequence of rice genomes, functional proteomics and the mining of protein expression profiles certainly constitute the next challenge for coming years in rice proteomics. The rice proteome analysis data provides necessary information to the breeders and greatly helps in understanding the growth and defense mechanism of the plant.

Acknowledgements

This work was supported by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences and MAFF Rice Genome Project PR.

References

Bae, M.S., Cho, E.J., Choi, E.-Y., Park, O.K., 2003. Analysis of the *Arabidopsis* nuclear proteome and its response to cold stress. Plant J. 36, 652–663.

Claros, M.G., Vincens, P., 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. 241, 779–786.

Cleveland, D.W., Fischer, S.G., Kirschner, M.W., Laemmli, U.K., 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J. Biol. Chem. 252, 1102–1106.

Devos, M.K., Gale, D.M., 2000. Genome relationships: the grass model in current research. Plant Cell 12, 637–646.

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., et al., 2002. Sequence and analysis of rice chromosome 4. Nature 420, 316–320.

Folta, K.M, Kaufman, L.S., 2000. Preparation of transcriptionally active nuclei from etiolated *Arabidopsis thaliana*. Plant Cell Rep. 19, 504–510.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., et al., 2002. A draft sequence of rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296, 92–100.

Gomez, S.M., Bil, K.Y., Aguilera, R., Nishio, J.N., Faull, K.F., Whitelegge, J.P., 2003. Transit peptide cleavage sites of integral thylakoid membrane proteins. Mol. Cell. Proteomics 2, 1068–1085.

Heazlewood, J.L., Howell, K.A., Whelan, J., Millar, A.H., 2003. Towards an analysis of the rice mitochondrial proteome. Plant Physiol. 132, 230–242.

Hirano, H., 1997. Screening of rice genes from the cDNA catalog using the data obtained by protein sequencing. J. Protein Chem. 16, 533–536.

Hirano, H., Kawasaki, H., Sassa, H., 2000. Two-dimensional gel electrophoresis using immobilized pH gradient tube gels. Electrophoresis 21, 440–445.

Imin, N., Kerim, T., Weinman, J.J., Rolfe, B.G., 2001. Characterization of rice anther proteins expressed at the young microspore stage. Proteomics 1, 1149–1161.

Kim, S.T., Cho, K.S., Jang, Y.S., Kang, K.Y., 2001. Two-dimensional electrophoretic analysis of rice proteins by polyethylene glyco fractionation for protein arrays. Electrophoresis 22, 2103–2109.

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

Komatsu, S., Kajiwara, H., Hirano, H., 1993. A rice protein library: a data-file of rice proteins separated by two-dimensional electrophoresis. Theor. Appl. Genet. 86, 935–942.

Komatsu, S., Muhammad, A., Rakwal, R., 1999a. Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. Electrophoresis 20, 630–636.

Komatsu, S., Rakwal, R., Li, Z., 1999b. Separation and characterization of proteins in rice (*Oryza sativa*) suspension cultured cells. Plant Cell Tiss. Org. Culture 55, 183–192.

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

Lamond, A.I., Earnshaw, W.C., 1998. Structure and function in the nucleus. Science 280, 547–553.

Lockhart, J.D., Winzeler, A.E., 2000. Genomics, gene expression and DNA arrays. Nature 405, 827–835.

Mikami, S., Hori, H., Mitsui, T., 2001. Separation of distinct components of rice golgi complex by sucrose density gradient centrifugation. Plant Sci. 161, 665–675.

Mikami, S., Kishimoto, T., Hori, H., Mitsui, T., 2002. Technical improvement to 2D-PAGE of rice organelle membrane proteins. Biosci. Biotechnol. Biochem. 66, 1170–1173.

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

Morre, D.J., Andersson, B., 1994. Isolation of all major organelles and membranous cell components from a single homogenate of green leaves. Meth. Enzymol. 228, 412–419.

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bioassay with tobacco tissue culture. Physiol. Plant 15, 473–497.

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

Pandey, A., Mann, M., 2000. Proteomics to study genes and genomics. Nature 405, 837–845.

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

Peltier, J.B., Frisco, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., van Wijk, K.J., 2000. Proteomics of the chloroplast: systematic identification and targeting analysis of luminal and peripheral thylakoid proteins. Plant Cell 12, 319–342.

Sasaki, T., Burr, B., 2000. International rice genome sequencing project: the effort to completely sequence the rice genome. Curr. Opin. Plant Biol. 3, 138–141.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., et al., 2002. The genome sequence and structure of rice chromosome 1. Nature 420, 312–316.

Shen, S., Matsubae, M., Takao, T., Tanaka, N., Komatsu, S., 2002. A proteomic analysis of leaf sheath from rice. J. Biochem. 132, 613–620.

Tanaka, N., Konoshi, H., Khan, M.M.K., Komatsu, S., 2004a. Proteome analysis of rice tissues by two-dimensional electrophoresis: an approach to the investigation of gibberellin regulated proteins. Mol. Gen. Genomics 270, 485–496.

Tanaka, N., Fujita, M., Handa, H., Murayama, S., Uemura, M., Kawamura, Y., Mitsui, T., Mikami, S., Tozawa, Y., Yoshinaga, T., Komatsu, S., 2004b. Proteomics of rice cell: systematic identification of the protein population in subcellular compartments. Mol. Gen. Genomics (in press).

Tsugita, A., Kamo, M., Kawakami, T., Ohki, Y., 1996. Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. Electrophoresis 17, 855–865.

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

Woo, S.H., Fukuda, M., Islam, N., Takaoka, M., Kawasaki, H., Hirano, H., 2002. Efficient peptide mapping and its application to identify embryo proteins in rice proteome analysis. Electrophoresis 23, 647–654.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79–92.

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

Md. Monowar Karim Khan was born in Kishoreganj, Bangladesh. He obtained his B.Sc. Ag (Hons) and M.Sc. (Ag) in Soil Science from the Bangladesh Agricultural University in 1983 and 1984, respectively. He was employed at the Bangladesh Institute of Nuclear Agriculture in 1986. In 1991, he was awarded Japanese Government Scholarship (Monbusho) and obtained his Ph.D. from the Chiba University, Japan in 1995. His Ph.D. Thesis work was focused mainly on nitrogen fixation in peanut. From 1997 to 1999 he was awarded Science and Technology Agency Postdoctoral Fellowship and pursued research on nodulation mechanism in legumes using reporter gene. Since 2001, he has been employed as a postdoctoral researcher in Setsuko Komatsu's laboratory at the National Institute of Agrobiological Sciences, Japan. His main research interests are nitrogen fixation in legumes and non-legumes, proteomics and molecular mechanism of signal transduction in cell.

Setsuko Komatsu obtained her Ph.D. from Meiji Pharmaceutical University, Japan. Her Ph.D. Thesis work was focused on the role of protein kinase dependent phosphorylation during fertilization of mammalian. She employed at Meiji Pharmaceutical University in 1980, and then also employed at Keio University, School of Medicine in 1985. Since 1990, she employed as senior researcher at the Department of Molecular Biology in National Institute of Agrobiological Sciences, and she initiated work on rice proteomics. In 2000, she became the head of Laboratory of Gene Regulation in National Institute of Agrobiological Sciences. Her main research interests are within the field of proteomics and molecular biology with a special focus on signal transduction in cell.