

Structural homology between semidwarfism-related proteins and glutelin seed protein in rice *(Oryza sativa* **L.)**

H. Hirano 1,2, S. Komatsu 1, A. Nakamura 3, E Kikuchi 3, H. Kajiwara 1, S. Tsunasawa 2 and E Sakiyama 2

1 Department of Molecular Biology, National Institute of Agrobiological Resources, Kannondai, Tsukuba, Ibaraki 305, Japan

2 Institute for Protein Research, University of Osaka, Suita, Osaka 565, Japan

3 Institute of Agriculture and Forestry, University of Tsukuba, Tsukuba 305, Japan

Received February 6, 1991; Accepted April 30, 1991 Communicated by G.S. Khush

Summary. Two semidwarfism-related proteins, SRP-1 and SRP-2, were detected as major spots in a long-culm rice cultivar, Norin 29 and its semidwarf near-isogenic line, SC-TN1, respectively, by two-dimensional gel dectrophoresis (2D-PAGE). The testcross showed that SRP-I and SRP-2 are controlled by codominant alleles, *Srp-1* and *Srp-2,* respectively, at a single locus *Srp.* This locus was considered to be closely linked with the semidwarfing gene locus *sd-1.* SRP-1 and SRP-2 were separated by 2D-PAGE, electroblotted onto a polyvinylidene difluoride membrane, and sequenced by a gas-phase protein sequencer. The N-terminal amino acid sequences, however, could not be determined due to the blockage of the N-terminals of these proteins. After removal of the N-terminal residue with pyroglutamyl peptidase given to the membrane, the amino acid sequence in the N-terminal region was determined. The N-terminal and internal amino acid sequences of SRP-I and SRP-2 were highly homologous with those of the glutelin α -subunits of seed endosperm storage protein, which were deduced by the eDNA sequences. In the seed endosperms of Norin 29 and SC-TN1, a total of eight glutelin α -subunits was identified by 2D-PAGE. The amino acid sequences in the N-terminal and internal regions of these proteins were determined. This experiment confirmed that SRP-1 and SRP-2 are almost identical in structure with the glutelin α_{5a} - and α_{5b} -subunits, respectively, which were identified in several organs such as endosperms, embryos, and leaves, unlike the other glutelin α -subunits.

Key words: Rice - Semidwarf near-isogenic line - Semi d warfism-related proteins – Glutelin – Pyroglutamic acid

Introduction

One of the most important objectives in rice breeding has been to breed semidwarf cultivars. To date, many semidwarf cultivars have been produced by conventional cross-breeding. The semidwarfing gene most frequently introduced in the course of breeding is a single recessive gene, *sd-I,* in Dee-geo-woo-gen (Chang et al. 1965; Hargrove et al. 1980; Kikuchi et al. 1985).

In a previous study (Nakamura et al. 1991a), we identified two semidwarfism-related proteins, SRP-I and SRP-2, which appear specifically as major polypeptides in the seed embryos of a tall Japanese cultivar, Norin 29, and its semidwarf near-isogenic line, SC-TN1, which carries *sd-l.* The results of testcrosses showed that SRP-1 and SRP-2 are controlled by codominant alleles, *Srp-1* and *Srp-2,* respectively, at a single locus *Srp.* Recently, we observed that the *Srp* locus is closely linked with the *sd- 1* locus at a recombination value of 15.5 % (Nakamura et al. 1991 b). Therefore, we considered that the isolation of the *Srp* gene may make it possible to identify the *sd-1* gene by gene manipulation techniques such as gene walking or jumping from *Srp.*

In the present study, in order to isolate the *Srp* gene, we analyzed the amino acid sequences of the SR proteins. The results of a previous experiment suggested that the N-terminal amino group of these proteins was blocked (Nakamura et al. 1991 a), indicating that the N-terminal sequence analysis by Edman degradation could not be performed without the removal of the blocked N-terminal residue. Therefore, in this study, we attempted to remove the N-terminal residue from SR proteins to determine the amino acid sequence in the N-terminal region. To obtain further detailed information on the sequence, we examined the internal amino acid sequences of these proteins. The results revealed that SR proteins

are structurally homologous with the glutelin α -subunits from the rice seed endosperms.

Materials and methods

Plant materials

A rice *(Oryza sativa* L.) cultivar, Norin 29, and its semidwarf near-isogenic line, SC-TN1, were used in this study. SC-TN1 had been obtained by transferring a semidwarf gene *(sd-1)* from the cultivar Taichung Native 1 into Norin 29 through seven backcrosses (Kikuchi et al. 1985; Nakamura et al. 1991a).

Two-dimensional gel electrophoresis

A portion $(100 \mu g)$ of embryos or endosperms from the dry mature seeds was removed, homogenized with $10 \mu l$ of lysis buffer (O'Farrell 1975), and centrifuged at $15,000 \times g$ for 5 min. The supernatant was subjected to two-dimensional gel electropboresis (2D-PAGE) as described by Hirano (1982).

N-terminal amino acid sequence analysis

After separation by 2D-PAGE, the proteins were electroblotted onto a polyvinylidene difluoride (PVDF) membrane (ProBlott, Applied Biosystems, Foster City) and detected by Coomassie blue staining (Hirano 1989). The portions of the PVDF membrane carrying the electroblotted protein were cut out and mounted into the upper glass block of the reaction chamber in a gas-phase protein sequencer (477A, Applied Biosystems). Edman degradation was performed according to the standard program supplied by Applied Biosystems. The released phenylthiohydantoin (PTH) amino acid derivatives were identified by the on-line system of high-performance liquid chromatography (120A, Applied Biosystems).

Removal of blocked N-terminal amino acid

After electroblotting of the blocked proteins, the PVDF membrane was soaked in 100 mM acetic acid containing 0.5% w/v polyvinylpyrrolidone-40 at 37 $^{\circ}\mathrm{C}$ for 30 min. The membrane was washed with deionized water ten times and soaked in 0.1 M phosphate buffer (pH 8.0) containing 5 m dithiothreitol and 10 mM ethylenediaminetetraethylacetic acid. After addition of 5 gg pyroglutamyl peptidase (Boehringer Mannheim, Mannheim), the reaction solution was incubated at 30° C for 24 h. The membrane was washed with deionized water, dried, and applied to the gas-phase protein sequencer.

Internal amino acid sequence analysis

The proteins were separated by 2D-PAGE and stained with Coomassie blue. The gel pieces containing the protein were removed and the protein was electroeluted from the gel pieces. The electroelution was performed in 0.375 M Tris-glycine buffer (pH 8.8) containing 1% w/v sodium dodecyl sulfate (SDS), using the Electrophoretic Concentrator (M1750, ISCO, Lincoln) at 2 W constant power for 2 h. After the electroelution, the protein solution was dialyzed against deionized water overnight, lyophilized, and dissolved in 20 μ 1 0.125 M Tris-HCl buffer (pH 6.8) containing 0.1% w/v SDS and 20% v/v glycerol. Ten microliters of *Staphylococcus aureus* V8 protease (Pierce, Rockford) $(0.1-0.2 \mu g/\mu l)$ in deionized water was overlaid with 10 μl 0.125 M Tris-HCl buffer (pH 6.8) containing 0.1% w/v SDS, 20% v/v glycerol, and 0.001% w/v Bromophenol blue. Electrophoresis was performed until the sample was stacked in the upper gel and then interrupted for 30 min to digest the protein

(Cleveland 1977). Electrophoresis was continued and the separated digests were electroblotted onto the PVDF membrane and sequenced by the gas-phase protein sequencer (Hirano and Watanabe 1990).

Homology search of amino acid sequences

The amino acid sequences obtained were compared with sequences of over 16,000 proteins contained in the amino acid sequence data base SEQDB (Protein Research Foundation, Osaka).

Quantitative changes of embryo and endosperm proteins during germination

Mature seeds of rice were immersed in water for 1, 3, 12, and 16 days at 25 °C. The embryos and endosperms were removed from the seeds. The proteins were then separated by 2D-PAGE and stained with Coomassie blue to measure their relative volume.

Detection of glycoproteins

Proteins from the seed endosperms were separated by 2D-PAGE, electroblotted onto a PVDF membrane, and reacted with peroxidase-coupled concanavalin A according to the procedure described by Kijimoto-Ochiai et al. (1985).

Results

Amino acid sequences of SR proteins

Proteins from the seed embryos of Norin 29 and SC-TNI were separated by 2D-PAGE (Fig. 1) and electroblotted from the gel onto the PVDF membrane. Among the electroblotted proteins, SRP-I and SRP-2 were sequenced by the gas-phase sequencer. However, no PTHamino acid was released, suggesting that the N-terminals of SRP-1 and SRP-2 were blocked, which is consistent with the previous results (Nakamura et al. 1991 a).

The electroblotted SRP-1 and SRP-2 were digested on the PVDF membrane with pyroglutamyl peptidase and subjected to the N-terminal amino acid sequence analysis. The pyroglutamyl peptidase digestion allowed the Edman degradation of SRP-I and SRP-2. The sequences of 16 and 9 residues from the N-terminals of SRP-1 and SRP-2, respectively, were determined (Fig. 2). The N-terminal pyroglutamic acid of SRP-1 and SRP-2 was considered to be formed by in vivo cyclization of glutamine or glutamic acid.

The embryo proteins were separated by 2D-PAGE, and SRP-I and SRP-2 were separately electroeluted from the gel. After the electroelution, SRP-I and SRP-2 were digested with *S. aureus* V8 protease on the SDS-PAGE gel. The digests were separated by SDS-PAGE and electroblotted from the gel onto the PVDF membrane to determine the internal amino acid sequences (Fig. 2).

Search for structural homology

The structural homology search using the SEQDB data base showed that the amino acid sequences of SRP-1 and

Fig. 1 A and B. 2D-PAGE patterns of SRP-1 and SRP-2. The adjacent proteins (a1-a3) were identified as not being glutelin subunits based on the analysis of their N-terminal and internal amino acid sequences. A Norin 29; B semidwarf near-isogenic line SC-TN1. Right to left: isoelectric focusing for first dimension; up to down: SDS-PAGE for second dimension. Detected by Coomassie blue staining

Fig. 2. Amino acid sequences of the seed glutelin α -subunits. 1)-4) The sequences deduced from nucleotide sequences. 1) Takaiwa et al. (1989); 2) Masumura et al. (1989); 3) and 4) Takaiwa et al. (1987). The partial sequences were obtained by actual protein sequencing of the endosperm glutelin α_1 -, α_2 -, α_3 -, α_4 -, α_{5a} -, and α_{5b} -subunits and the embryo glutelin SRP-1 and SRP-2. The different residues between 1) and 2), and between 3) and 4) are indicated by $dots. < Q$ represents pyroglutamic acid

SRP-2 were highly homologous with that deduced from the nucleotide sequence of type I cDNA (Takaiwa et al. 1987) encoding glutelin, which is a major storage protein in the seed endosperms. In the regions of SRP-I or SRP-2 examined, the sequences were identical with that deduced from type I cDNA, except that the 60th residue was valine, while the corresponding residue in SRP-I and SRP-2 was alanine, and the 122nd residue was glutamine, while it was histidine in SRP-2.

Identification of the endosperm glutelin subunits

In the seed endosperms of Norin 29 and SC-TN1, a total of eight glutelin a-subunits was detected by 2D-PAGE. To determine which glutelin subunits were identical or most homologous with SRP-I and SRP-2, the structural differences of the eight glutelin subunits were analyzed.

All these glutelin subunits were found to have blocked N-termini. Therefore, after 2D-PAGE, the glutelin subunits electroblotted onto the PVDF membrane were digested with pyroglutamyl peptidase to remove the blocked N-terminal amino acid. After removal, the amino acid sequences of these protein subunits could be determined in the N-terminal regions (Fig. 2).

The glutelin fragments obtained after in situ protease digestion on SDS-PAGE gel were electroblotted onto the PVDF membrane and their N-terminal amino acid sequences were determined.

The results of these experiments revealed that the amino acid sequences of SRP-I and SRP-2 were most homologous with those of the glutelin α_{5a} - and α_{5b} -subunits.

As described previously (Nakamura et al. 1991 a), SRP-1 and SRP-2 were detected as proteins with different isoelectric points, 7.4 and 7.7, respectively, but with an identical molecular weight, 32 kDa, by 2D-PAGE. In this study, we examined whether or not the endosperm glutelin α -subunits with different isoelectric points between Norin 29 and SC-TN1 were present, and found that the glutelin α_{5a} - and α_{5b} -subunits showed the same electrophoretic pattern in isoelectric focusing as SRP-1 and SRP-2, respectively (Fig. 3).

To confirm the difference of these polypeptides, we mixed the endosperm and embryo protein solutions, and separated the proteins by 2D-PAGE. The electrophoretic mobility of SRP-I and SRP-2 was identical with that of the glutelin α_{5a} - and α_{5b} -subunits, respectively.

From these results, we concluded that SRP-1 and SRP-2 are identical with glutelin α_{5a} - and α_{5b} -subunits, respectively (Table 1).

SPR-1 and SRP-2 are glycoproteins (Nakamura et al. 1991 a). The electroblotted endosperm glutelin α -subunits were examined by the concanavalin A peroxidase method to determine whether or not they were glycoproteins.

Fig. 3A-C. 2D-PAGE patterns of the seed endosperm glutelin α -subunits. A Cultivar Norin 29; **B** A + C mixture; C semidwarf near-isogenic line SC-TN1. *Right* to *left:* isoelectric focusing for first dimension; *up* to *down:* SDS-PAGE for second dimension. Detected by Coomassie blue staining

It was suggested that the glutelin α_{5a} - and α_{5b} -subunits were glycosylated, as well as the glutelin α_1 -, α_2 -, α_3 -, and α_4 -subunits, but it could not be ascertained whether or not the glutelin α_6 - and α_7 -subunits were glycosylated.

Complementary DNA encoding the glutelin subunits

The amino acid sequences of the glutelin α -subunits were compared with those deduced from the nucleotide sequences of DNAs encoding the glutelin subunits. This comparison showed that the sequence of glutelin α_1 was highly homologous with those deduced from the nucleotide sequence of cDNA as reported by Masumura et al. (1989). The amino acid sequences of the glutelin α_2 -, α_3 -, and α_4 -subunits were found to be identical with that deduced from the nucleotide sequences of type II cDNA cloned by Takaiwa et al. (1987). The sequences of glutelin α_{5a} (SRP-1) and α_{5b} (SRP-2) were homologous with that deduced from the sequence of type I cDNA as cloned by Takaiwa et al. (1987). No genes encoding glutelin α_6 and α ₇ have been reported so far. In these subunits, the N-ter-

Glutelin subunit	pI	Mr (kDa)	Embryo		Endosperm		Glvco- sylation	Gene	Locus	cDNA ^a	Identified differ- ence in sequence
			Norin 29	SC-TN1		Norin 29 SC-TN1					
α_1	8.0	31				$^{+}$	┿				None
α_{2}	7.8	31				$^{+}$	\div			$2)$ (Type II)	None
α_3	7.7	32				$^{+}$	$+$			$2)$ (Type II)	None
α_4	7.7	31				$^{+}$	$+$			$2)$ (Type II)	None
α_{5a}	7.4	32	$+(SRP-1)$		\pm		$^+$	$Srp-1$	Srp	$2)$ (Type I)	60th Val \rightarrow Ala
α_{5b}	7.7	32		$+$ (SRP-2)			$^{+}$	$Srp-2$	Srp	$2)$ (Type I)	122nd Gln \rightarrow His
α_6	7.3	31			\pm	+				Unknown	
α_{7}	7.2	32				\div				Unknown	

Table 1. Characteristics of eight glutelin α -subunits in rice

^a In these experiments, mRNAs were extracted from cultivar Nipponbare. We found that there is no difference in the glutelin subunit compositions and the the actual amino acid sequences between Norin 29 and Nipponbare. Reference 1) Masumura et al. 1989; 2) Takaiwa et al. 1987

minal glutamine that is present in the other subunits was deleted. The α_6 - and α_7 -subunits were considered to be encoded by the gene that is structurally similar to the type I and II cDNAs, but not identical.

Glutelin species in the embryos

As described above, we found that the proteins identical with the endosperm glutelin α -subunits appeared in the embryos. It remained to be determined whether only the glutelin α_{5a} - and α_{5b} -subunits or all the glutelin α -subunits were expressed in the embryos.

We separated the embryo proteins by 2D-PAGE and analyzed the amino acid sequences of the proteins that showed a similar molecular weight and isoelectric point to those of the endosperm glutelin α -subunits. However, no glutelin α -subunits except for the α_{5a} - and α_{5b} -subunits could be identified in the embryos of Norin 29 and SC-TN1 (Fig. 1).

Quantitative changes of glutelin subunits in embryos and endosperms during germination

In the embryos, the amount of the glutelin α_{5a} - and α_{5b} subunits increased during germination (Nakamura et al. 1991 a). However, in the endosperms, the amount of all the glutelin α -subunits gradually decreased during germination, and they could no longer be detected 16 days after water immersion.

Discussion

In the previous report (Nakamura et al. 1991 a), we suggested that SR proteins had blocked N-termini. We developed a simple technique to remove the blocked N-terminal residue using pyroglutamyl peptidase on a PVDF membrane by modification of the conventional methods (Podell and Abraham 1978; Moyer et al.

1990). This technique allowed us to deblock and subsequently to sequence subnanomole amounts of protein on the PVDF membrane. In the present study, using this technique, we were able to determine the N-terminal amino acid sequences of SR proteins and glutelin α -subunits which were separated by 2D-PAGE.

There are eight glutelin species in the endosperms of Norin 29 and SC-TN1. Among them, the glutelin α_{5a} and α_{5b} -subunits are expressed not only in the endosperms but also in the embryos and young leaves. Although the amount of embryo glutelin α -subunits increased during germination as described in the previous paper (Nakamura et al. 1991 a), the amount of all the endosperm glutelin α -subunits decreased. This result suggests that the mechanism regulating the expression of the genes encoding the same glutelin subunits should be different between the embryos and endosperms. The glutelin is thought to act as a source of nitrogen and amino acids for germination in the endosperms. However, the glutelin α_{5a} - and α_{5b} -subunits in the embryos are likely to have a different function from the storage proteins.

The endosperm glutelin is synthesized as a precursor polypeptide consisting of a signal peptide, α - and β -subunits. The precursor polypeptide is processed cotranslationally and posttranslationally to generate disulfidebonded α - and β -subunits (Yamagata et al. 1982). We separated the embryo glutelin α_{5a} - and α_{5b} -subunits by 2D-PAGE under nonreducing conditions (data not shown). Under such conditions, the α_{5a} - and α_{5b} -subunits were not identified in the positions where they would be detected under reducing conditions as shown in Fig. 1. These results indicate that the embryo glutelin α_{5a} - and α_{5b} -subunits are linked with another polypeptide, possibly glutelin β -subunit, via disulfide bonding like the endosperm glutelin subunits.

In the previous report (Nakamura et al. 1991 a), we reported that SRP-1 and SRP-2 are glycoproteins that can be identified by the concanavalin A peroxidase method. We examined whether or not endosperm glutelin α -subunits contained sugar chains by the same method. The results suggested that the glutelin α_1 -, α_2 -, α_3 -, and α_4 -subunits in the endosperms, and the glutelin α_{5a} - and α_{5b} -subunits in both embryos and endosperms, were glycosylated. Although the N-linked oligosaccharides are

thought to be detected by this concanavalin A peroxidase method, the amino acid sequences of all published glutelin α -subunits lack the typical tripeptide Asn-X-Thr or Asn-X-Ser signal for N-glycosylation. The structures of oligosaccharides in the glutelin α -subunits remain to be determined.

The *Srp* locus is linked with the *sd-1* locus the loci identified so far in rice, with a recombination value of 15.5% (unpublished data). The size of the rice genome is estimated at approximately $10⁶$ kbp. Accordingly, the value of 15.5% roughly corresponds to $500-1,000$ kbp. Since the cDNAs and genes encoding the glutelin have been cloned, it may become possible to isolate the *sd-1* gene in the future by applying the gene walking or jumping technique.

References

- Chang TT, Morishima H, Huang CS, Tagumpay T, Tateno K (1965) Genetic analysis of plant height, maturing and other quantitative traits in the cross of Peta x I-geo-tze. J Agric Assoc China News Serv 39:1-8
- Cleveland DW, Fischer SG, Kirschner MW, Laemmli UK (1977) Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. J Biol Chem 252: t 102-1106
- Hargrove TH, Coffman WR, Cabanilla VL (1980) Ancestry of improved cultivars as parents in Asian rice breeding programs. Crop Sci 20:721-727

Note added in proof. Although the codominant alleles controlling the SR proteins were defined here as *Srp-1* and *Srp-2* genes at *Srp* locus, they should be tentatively referred to as *Srpl(t)* and *Srp2(t)* genes at *Srp(t)* locus, according to the recommendation of International Rice Commission (See Nakamura et al. 1991 b).

- Hirano H (1982) Varietal differences of leaf protein profiles in mulberry. Phytochemistry 21:1513-1518
- Hirano H (1989) Microsequence analysis of winged bean seed proteins electroblotted from two-dimensional gel. J Protein Chem 8:115-130
- Hirano H, Watanabe T (1990) Microsequencing of proteins electrotransferred onto immobilizing matrices from polyacrylamide gel electrophoresis: application to an insoluble protein. Electrophoresis 11:573-580
- Kijimoto-Ochiai S, Katagiri YU, Ochiai H (1985) Analysis of N-linked oligosaccharide chains of glycoproteins on nitrocellulose sheets using lectin-peroxidase reagents. Anal Biochem 147:222-229
- Kikuchi F, Itakura N, Ikehashi H, Yokoo M, Nakane A, Maruyama K (1985) Genetic analysis of semidwarfism in high-yielding rice varieties in Japan. Bull Natl Inst Agric Sci Ser D 36:125-145
- Masumura T, Kidzu K, Sugiyama Y, Mitsukawa N, Hibino T, Tanaka K, Fujii S (1989) Nucleotide sequence of a cDNA encoding a major rice glutelin. Plant Mol Biol 12:723-725
- Moyer M, Harper A, Payne G, Ryals J, Fowler D (1990) In situ digestion with pyroglutamate aminopeptidase for N-terminal sequencing electroblotted proteins. J Protein Chem 9:282-283
- Nakamura A, Hirano H, Kikuchi F (1991 a) Identification and genetic analysis of semidwarfism-related proteins in rice *(Oryza sativa* L.). Theor Appl Genet 81: 376- 380
- Nakamura A, Hirano H, Kikuchi F (1991 b) A protein gene *Srp(t)* linked with a semidwarfing gene *sd-1* in rice *(Oryza sativa* L.). Japan J Breed 41:517-521
- O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007-4021
- Podell DN, Abraham GN (1978) A technique for the removal of pyroglutamic acid from the amino terminus of proteins using calf liver pyroglutamate aminopeptidase. Biochem Biophys Res Commun 81:176-185
- Takaiwa F, Kikuchi S, Oono K (1987) A rice glutelin family A major type of glutelin mRNAs can be divided into two classes. Mol Gen Genet 208:15-22
- Takaiwa F, Kikuchi S, Oono K (1989) The complete nucleotide sequence of new type cDNA coding for rice storage protein glutelin. Nucleic Acids Res 17:3289
- Yamagata H, Sugimoto T, Tanaka K, Kasai Z (1982) Biosynthesis of storage proteins in developing rice seeds. Plant Physiol 70:1094-1100