Bot. Bull. Acad. Sin. (1994) 35: 9 – 14

(Invited review paper)

On the molecular biology of rice glutelin[1]

Hsin-Kan Wu

Institute of Botany, Academia Sinica, Taipei, Taiwan 115, Republic of China

Abstract. Rice grains contain 8% to 18% protein, mostly storage type, which can be classified as water soluble salt soluble, ethanol soluble, and acid- or alkali- soluble; the alkali soluble protein is usually referred to as glutelin. During development of the rice caryopsis, glutelin starts to synthesize at 5 days after flowering (DAF) and reaches its plateau at 10 DAF at gathered polysomes and/or membrane bound ribosomes. Glutelin deposits in a protein body in the endosperm cells, especially those cells located in the outmost layers. Purification of glutelin can be accomplished by sequential extractions. The preglutelin is a 57 kiloDalton (kDa) peptide in which a signal peptide sequence is located at the N-terminal. After cotranslational processing, two subunits, α and β, are produced which can be resolved by sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) into two groups of bands around 33 kDa and 22 kDa, respectively. With iso-electric focusing (IEF), α and β subunits can be separated into bands appearing in acidic and basic regions, respectively. In both SDS-PAGE and IEF there is a difference in the distribution pattern of the component bands of glutelin from cultivated and from wild rices and also from among the wild species. According to the sequence of the 5' regions located upstream of its coding sequence, the glutelin gene family is composed of at least three member genes, each of which has a 5' region of 0.5 kb, 0.9 kb, or 1.2 kb. It is not known whether the directed repeats found in the 5' regions are closely related to the enhancement of glutelin gene expression, though several protein binding elements have been reported. In the sequence down stream of the coding sequence, no vacuolar targeting signal, such as CTPP, nor a sequence similar to CTPP has been identified. Genetic engineering of glutelin is possible but requires collaboration among disciplines.

Keywords: Expression; Genetic engineering; Glutelin; Rice; Separation; Structure; Synthesis.

Contents

Introduction

It is known that the protein content of rice varies from 8% to 18% among varieties (Tong et al., 1970). The proteins are stored mostly in the seed's endosperm, and are referred to as storage proteins. They can be grouped into three categories based on their solubility, namely water soluble salt soluble, ethanol soluble, and alkali- or acid- soluble. Glutelin (alkali soluble) is the major storage protein, comprising more than 80% of the total (Lin et al., 1971). Cytologically, it has been shown that rice storage proteins are deposited as a protein body, which can be found in the outermost cell layers of the endosperm (Mitsuda et al., 1967; Wu and Chen,1978). Three types of rice protein bodies have been reported, namely densely stained sphere, less densely stained sphere, and angular (Bechtel et al., 1978; Bechtel and Juliano, 1980). We classified them into only two types:

[1] This paper was presented at the Symposium on Botany (December 7 – 22, 1992) in commemoration of the 30th anniversary of the re-establishment in Taipei of the Institute of Botany, Academia Sinica. The abstract of the paper will also be published in the Institute of Botany, Academia Sinica Monograph Series No. 13 (1993).

round and angular. Furthermore, the number of protein bodies differs among varieties, with the developmental stages of 14, 21, and 28 days after flowering (DAF), and with the location (top, ventral, dorsal, and base) on the caryopsis (Harris and Juliano, 1978; Wu and Chen, 1978).

Site of Synthesis and Deposition of Glutelin

In my laboratory, the involvement of vesicles in the formation of storage protein bodies was suggested. The vesicles are coated with ribosomes and derive from the numerous Golgi apparati and/or endoplasmic reticula (ER) in the peripheral cells of an endosperm. As the content of the vesicles is not digestible by pepsin, the protein that is to be stored can not be synthesized and/or processed ahead of time in the Golgi apparati or ER, but is synthesized de novo on the surface of rough vesicles which partially cover a protein body core-nucleus. Formation of the core-nucleus can take place anywhere in the cytoplasmic matrix by the accretion of proteins synthesized from gathered polyribosomes. The formation of the angular protein body is similar to that of the round one because it appears in vacuoles which are usually surrounded by numerous vesicles (Wu and Chen, 1978). Bechtel and Juliano (1978) found that the round protein bodies are deposited within the rough endoplasmic reticulum (RER) and that the angular bodies are deposited via the Golgi apparati. Yamagata and Tanaka (1986) reported that the round ones (PB I) appear to be synthesized by membrane-bound polysomes or in the RER and pass through the RER membrane into its lumen where they aggregate and deposit. The angular ones (PB II) seem to be synthesized by membrane-bound polysomes as a precursor protein and to become sequestered in the cisternal space of the RER, from where they are transferred to the vacuole. It may be concluded that rice storage protein is synthesized at gathered polysomes and/or membrane-bound ribosomes, and is deposited in protein bodies, but there is no agreement about the route of transport between the sites of synthesis and deposition.

Immunocytochemical studies by Krishnan et al. (1986) showed that glutelins are present in the angular protein bodies, while prolamines are present in both the large and small spherical bodies. This is in contradiction to the observations, made during in situ digestion experiments, that almost all of the round protein bodies can be simply dissolved by treating with 0.4% NaOH solution at 35 °C for 100 min. This means that the round protein bodies are alkali soluble glutelin. Furthermore, the number of round protein bodies exceeds that of the angular ones by 2 – 4 fold (Wu and Chen, 1978). The major storage protein, glutelin should have more room to be accommodated. The controversy may be solved by using monoclonal- instead of polyclonal- antibodies because the former are domain specific.

Separation of Glutelin Components

In 1982 Yamagata et al. reported that glutelin begins to accumulate about 5 days after flowering (DAF). It was resolved into two polypeptide groups, 22 to 23, and 37 to 39 kiloDaltons (kDa) by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of proteins from the starchy endosperm of rice (*Oryza sativa* L. Japonica cv. Koshihikari). In vitro protein synthesis also showed that the mRNAs coding for the two polypeptides groups are absent from developing caryopses and that the 57 kDa polypeptide is the major product. Thus, Yamagata et al. concluded that two subunits of glutelin are formed through post-translational cleavage of the 57 kDa polypeptides. This conclusion is consistent with the results of other workers (Luthe, 1983; Sarker et al., 1986; Krishnan et al., 1986; Shyur et al., 1988), although the molecular weights estimated by different authors are not the same. We have been able to separate glutelin by SDS-PAGE with finer resolution. On the gel, it appeared that both of the so-called 22 to 23 kDa and 37 to 39 kDa groups from Japonica and Indica cultivars and from wild species are composed of several bands with various molecular weights (Tarn, 1988). Isoelectric focusing in polyacrylamide gels showed that isoelectric points (pI) of the glutelin subunits (α and β polypeptides) range from 6.5 to 7.5 and from 9.4 to 10.3, respectively. At least twelve polypeptides of the α subunit and nine polypeptides of the β subunit could be separated by isoelectric focusing (Wen and Luthe, 1985). The basic subunit has been resolved sharply, however, only by nonequilibrium pH-gradient electrophoresis, which is able to separate approximately ten bands (Chen et al., 1987). In a recent isoelectric focusing analysis, we were able to separate glutelin into more than twelve bands. SDS-PAGE analysis shows the peptides of each pI value to be composed of one to three bands of different molecular weight. This suggests that rice glutelin is encoded by a gene family.

A more efficient and up to date method to separate glutelin may be that recently reported by Huebner et al. (1990). Using improved extraction and reverse phase high-performance liquid chromatography (RP - HPLC), they separated rice glutelin into about twenty major and minor peaks to differentiate rice varieties. RP-HPLC separation also revealed that the quantity of glutelin differs among varieties. If the peaks could be qualitatively examined by running an SDS-PAGE, this technique would be more applicable in rice breeding for better glutelin. It would be also interesting to use this separation technique combined with peak investigation to compare the composition of the glutelins extracted

from wild rice species having different genomic backgrounds.

Structure of Glutelin Genes

As mentioned earlier in this review, the synthesis of rice glutelin is controlled by a gene family. The first isolation of glutelin cDNAs from cultivated rice was performed by Tanaka and Ogawa (1985). The ^{32}P-labeled single-strand cDNA probe they used to fish glutelin cDNAs from a cDNA library was reverse-transcribed from the poly(A)$^+$ RNA that was extracted from the rough endoplasmic reticulum of rice caryopses. The isolated cDNA (pREE61) was soon sequenced (Takaiwa et al., 1986). Later, more cDNA clones were isolated from a rice cDNA library and classified into two types. Those of type I are pREE61 and pREE103, those of type II are pREE77 and pREE99. Two polyadenylation signal sites were found in type I clones but only one site in type II (Takaiwa, Kikuchi, and Oono,1987). In my laboratory, restriction enzyme mapping and partial sequencing of our own glutelin cDNA clones (39 clones) enabled us to classify them into seven groups (Wu and Fu, 1988). The genomic structure of type II glutelin gene was determined by Takaiwa et al. (1987). Five glutelin genes from cultivated rice have been isolated by Takaiwa et al. (1988). All of these glutelin genes are interrupted by three short introns that have no extensive homology. Three other genomic clones for rice glutelin were isolated and studied by Okita et al. (1989). Comparison of DNA sequences from relevant regions of these clones showed that two of them, Gt1 and Gt2, are closely related. The 5' flanking regions and the coding sequences of these two clones have 84% and 91% homology, respectively. In contrast, another clone, Gt3, shows little or no homology to Gt1 and Gt3 in the 5' flanking sequences upstream of the putative CAAT boxes (-95 to -121) and exhibits significant divergence in all other portions of the gene. Sequencing is helpful in determining the structure of a gene. One should consider, however, to what extent of divergence two sequences can be regarded as two members of a gene family. Should the two flanking (5' and 3') regions be included when identifying genes?

The cloned type II glutelin genes have a 5' flanking region of about 0.9 kb (Takaiwa et al., 1987; Okita et al., 1989). When short DNA segments (25 or 29 mer) complementary to each end of the published 0.9 kb 5' region were synthesized and used as primers, 5' regions of several wild rice species were amplified by polymerase chain reaction. Consequently, we have obtained two new 5' regions in addition to the old 0.9 kb region from glutelin genes. One, 1.2 kb long, amplified from all the species tested, the other one, only 0.5 kb long, is from a species of CC genome. Homology between the 0.9 kb and 1.2 kb 5' region is relatively high, though both of them vary in length among glutelin genes of wild rice species. It seems that some short segments of the 1.2 kb region have been deleted to form the 0.9 kb 5' region, or some short segments have been added to the 0.9 kb region to form the 1.2 kb 5' region. In the case of the 0.5 kb 5' region, it was shown by sequencing that a big fragment of about 0.4 kb, beyond several short deletions, has been deleted. Another feature of this 5' region is that a fragment of 49 bp that is homologous to that of the 0.9kb 5' region is attached at its far upstream end, (Wu, Accession number M63124, GenBank, 1991). Structural changes in the 5' region of a gene may affect its expression. The three 5' regions isolated from glutelin genes of different wild rice species must have long been evolved because they have base deletions and additions as well as substitutions. It would be justifiable to postulate that they are the three member genes of the glutelin gene family.

Expression of Glutelin Genes

Until recently, there have been few studies of the glutelin gene expression. Remachandran and Raghavan (1990) observed a steady increase in the concentration of glutelin mRNA in pericarp and endosperm cells over a period of 6 to 18 days after anthesis, and the completed disappearance of this transcript by 24 days after anthesis. Northern blot hybridization of ^{32}P-labelled glutelin gene to rice grain poly(A)$^+$ RNA confirmed the temporal pattern of gene expression during the development of rice grain that was revealed by in situ hybridization using RNA probes. The 5' flanking region of type II glutelin gene (pREE 77) was examined for its regulatory function in transgenic tobacco (Takaiwa et al., 1991). It is specifically expressed in developing seeds, as opposed to leaves and stems. Deletion studies of the 5' region in tobacco indicated that the region between positions -441 and -237 is required for temporal and endosperm-specific expression of GUS activity. Kim and Wu (1990) have demonstrated that at least six protein-DNA complexes are formed between the 5' flanking region (-677 to -45) of glutelin gene pGL-5 and nuclear protein factors. Two of the protein-binding sequences contain the TGAGTCA motif, which is also present in the -300 element found in the 5' flanking regions of several storage protein genes of other crop plants. Further studies of the interaction of nuclear protein factors with the 5' region of rice glutelin gene were performed by Takaiwa and Oono (1990). They showed that the specific region between positions -272 and -99 of a type II glutelin gene(pHIR13-7) interacts with nuclear proteins extracted from immature seeds, but not with those from leaves and roots, and that the binding activity is closely correlated with the mRNA

levels of this gene during seed maturation. It may be summarized that in the 5' flanking region of glutelin genes there are several specific elemental sequences for binding proteins that are regulatory factors for gene expression, either spatially and/or temporally. It would be worthwhile to carry out experiments to assess the expression capabilities of the three 5' regions of glutelin genes isolated from wild rice species, by examining whether the naturally evolved gene structures comform to their gene functions.

Although both the carboxyl- and amino- terminal propeptide (CTPP and ATPP respectively) have been determined as vacuolar protein targeting determinants for barley lectin (Bednarek et al., 1991) this information is not yet available for rice glutelin.

Genetic Engineering of Glutelin

Based on current knowledge of the molecular biology of glutelin, it seems possible to engineer a more nutritionally balanced glutelin. This could be done by changing the amino acid composition of endogenous glutelin so that it contains high levels of certain desired amino acids, by introducing other seed protein genes which would correct the amino acid deficiency, and by increasing the level of expression of endogenous genes encoding proteins that have a more complete amino acid balance, as Larkins (1983) proposed for cereals. In practice, however, protein engineering needs to be guided by conventional genetic information, such as the fact that elevated lysine is inherited as a recessive gene and that increased lysine is correlated with decreased seed size. F_3 and F_4 data provide evidence for the transmission of high lysine genes to advanced germplasm in rice, as Schaeffer et al. (1989) reported. Schaeffer and Sharpe (1990) also found that selection for increased levels of lysine and threonine in rice also selects for a decrease in tryosine levels. The task of genetic engineering of glutelin, though possible, requires collaboration among geneticists, breeders, biochemists, and molecular biologists.

Literature Cited

Bechtel, D. B. and Y. Pomeranz. 1978. Ultrastructure of the mature ungerminated rice (*Oryza sativa*) caryopsis, the germ. Amer. J. Bot. **65:** 75 – 85.

Bechtel, D. B. and B. O. Juliano. 1980. Formation of protein bodies in the starchy endosperm of rice (*Oryza sativa* L.): A re-investigation. Ann. Bot. **45:** 503 – 509.

Bednarek, S.Y. and N. V. Raikhel. 1991. The barley lectin carboxyl-terminal propeptide is a vacu olar protein sorting determinant in plants. Plant Cell **3:** 1195 – 1206.

Chen, S. C. G., J. Chen, and L. Y. Hwang. 1987. Biochemical and immunological properties of glutelin from indica rice. Bot. Bull. Acad. Sin. **28:** 237 – 246.

Harris, N. and B.O. Juliano. 1977. Ultrastructure of endosperm protein bodies in developing rice grains differing in protein content. Ann. Bot. **41:** 1 – 5.

Huebner, F. R., J. A. Bietz, B.D. Webb, and B. O. Juliano. 1990. Rice cultivar identification by high-performance liquid chromatography of endosperm proteins. Cereal Chem. **67:** 129 – 135.

Kim, S. Y. and R. Wu. 1990. Multiple protein factors bind to a rice glutelin promoter region. Nucl. Acids Research **18:** 6845 – 6852.

Krishnan, H. B. and T. W. Okita. 1986. Structural relationship among the rice glutelin polypeptides. Plant Physiol. **81:** 748 – 753.

Krishnan, H. B., V. R. Franceschi, and T. W. Okita.1986. Immunochemical studies on the role of the Golgi complex in protein body formation in rice seeds. Planta **169:** 471 – 480.

Larkins, B. A. 1983. Genetic engineering of seed storage proteins. *In* T. Kosuge, C.P. Meredith, and A. Hollaender (eds.), Genetic Engineering of Plants – An Agricultural Perspective. Plenum Press, N.Y. 1983.

Lin, P. C., L. F. Lin, and S. T. Chow. 1971. Studies on the varietal characteristics of rice in quality. J. Taiwan Agr. Res. **20:** 31 – 41.

Luthe, D. S. 1983. Storage protein accumulation in developing rice (*Oryza sativa* L.) seeds. Plant Sci. Letters **32:** 147 – 158.

Mitsuda, H., K. Yasumoto, K. Murakami, T. Kusano, and H. Kishida. 1967. Protein bodies in rice endosperm. Memoirs No. 92, College of Agr. Kyoto Univ. Kyoto, Japan.

Okita, T. W., Y. S. Hwang, J. Hnilo, W. T. Kim, A. P. Aryan, R. Larson, and H. B. Krishnan. 1989. Structure and expression of the rice glutelin multigene family. J. Biol. Chem. **264:** 12573 – 12581.

Ramachandran, C, and V. Raghavan. 1990. Intracellular localization of glutelin mRNA during grain development in rice. J. Experimental Bot. **41:** 393 – 399.

Sarker, S. C., M. Ogawa, M. Takahashi, and K. Asada. 1986. The processing of a 57-kDa precursor peptide to subunits of rice glutelin. Plant Cell Physiol. **27:** 1579 – 1586.

Schaeffer, G. W., F. T. Sharpe, Jr. and J. T. Dudley. 1989. Segregation for endosperm lysine in F_2, F_3 and F_4 progeny from a cross of in vitro selected and unselected cultivars of rice. Theor. Appl. Genet. **77:** 176 – 183.

Schaeffer, G. W. and F. T. Sharpe, Jr. 1990. Modification of amino acid composition of endosperm proteins from in vitro selected high lysine mutants in rice. Theor. Appl. Genet. **80:** 841 – 846.

Shyur, L. F., K. K. Zia, and C. S. Chen. 1988. Purification and some properties of storage proteins in japonica rice. Bot. Bull. Acad. Sin. **29:** 113 – 122.

Takaiwa, F., S. Kikuchi, and K. Oono. 1986. The structure of rice storage protein glutelin precursor deduced from cDNA. FEBS Letters **206:** 33 – 35.

Takaiwa, F., S. Kikuche, and K. Oono. 1987. A rice glutelin gene family — a major type of glutelin mRNAs can be divided into two classes. Mol. Gen. Genet. **208:** 15 – 22.

Takaiwa, F., H. Ebinuma, S. Kikuchi, and K. Oono. 1987. Nucleotide sequence of a rice glutelin gene. FEBS Letters **221:** 43 – 47.

Takaiwa, F. and K. Oono. 1990. Interaction of an immature seed-specific trans-acting factor with the 5' upstream region of a rice glutelin gene. Mol. Gen. Genet. **224:** 289 – 293.

Takaiwa, F., K. Oono, and A. Kato. 1991. Analysis of the 5' flanking region responsible for the endosperm-specific expression of a rice glutelin chimeric gene in transgenic tobacco. Plant Mol. Biol. **16:** 49 – 58.

Tanaka, K. and M. Ogawa. 1985. Genetic analysis of rice storage proteins. Intl. Rice Genetics Symposium, 27 – 31 May 1985. Manila, Philippines.

Tarn, W.Y. 1988. A dissection for the composition of glutelin α subunit of *Oryza* species. Master thesis submitted to the College of Agr. Natl. Taiwan Univ. Taipei, ROC.

Tong, W. F., Y. E. Chu, and H. W. Li. 1970. Variation in protein and amino acid contents among genetic stock of rice. Bot. Bull. Acad. Sin. **11:** 55 – 60.

Wen, T. N. and D. S. Luthe. 1985. Biochemical characterization of rice glutelin. Plant Physiol. **78:** 172 – 177.

Wu, H. K. and J. T. Chen. 1978. Protein bodies in developing rice grain. Proc. Natl. Sci. Council. (ROC) **2:** 281 – 292.

Wu, H. K. and H. Y. Fu. 1988. Partial sequencing of rice glutelin clones. Abst. No. 202. The 2nd Intl. Congress of Plant Molecular Biology. Jerusalem, Nov. 13 – 18, 1988.

Wu, H. K. 1991. Accession no. M63124, GenBank.

Yamagata, H., T. Sugimoto, K. Tanaka, and Z. Kasai. 1982. Biosynthesis of storage proteins in developing rice seeds. Plant Physiol. **70:** 1094 – 1100.

Yamagata, H. and K. Tanaka. 1986. The site of synthesis and accumulation of rice storage proteins. Plant Cell Physiol. **27:** 135 – 145.

稻穀蛋白的分子生物學

吳信淦

中央研究院植物研究所

稻米含有的蛋白在 8-18% 之間，主要由貯存蛋白所貢獻。貯存蛋白依溶性可分成四種，即水溶性、鹽溶性、醇溶性及鹼溶性。後者稱爲穀蛋白，佔貯存蛋白總量的 80% 以上。貯存蛋白分佈於稻穎果的胚乳，以顆粒形態存在於胚乳的最外幾層細胞內。分層萃取法可將四種溶性蛋白一一純化之。栽培稻穎果內穀蛋白的合成與積聚於開花後五天開始，十天時最爲旺盛。聚核糖體及帶有核糖體的膜系構造，如粗面泡囊，爲其合成位置。野生稻的情況則因物種而異。穎果內合成的穀蛋白爲一57kD包括信號片段的前驅分子，經加工成爲 α 及 β 兩個次元分子，各在SDS—PAGE膠體的 33 kD 及 22 kD 處順次顯示一環帶群。等電點集焦膠體上，此兩次元分子則各可順次分爲若干酸性及鹼性環帶。栽培稻與野生稻間以及野生稻諸物種間穀蛋白組成環帶在此兩種膠體上比較時，其分布均不相同。穀蛋白 cDNA 的核苷酸序列於1985 年初次出現，隨著有其整個基因的。根據穀蛋白基因轉錄區上游 5' 區的序列，此基因族至少由此 5' 區各爲 0.5kb、0.9kb及1.2kb的三種成員基因所組成。已知此 5' 區的若干片段與調控穀蛋白的結合有關；但同區內若干定向重複片段是否能加強穀蛋白基因的轉錄尚未得知。至穀蛋白基因轉錄區下游 3' 區外的序列是否有導向穀蛋白進入液泡的 CTPP 信號或與此信號相似的信號亦尚未見澄清。以遺傳工程改善穀蛋白的品質有其可能，但需各學門間的合作。

關鍵詞：稻；穀蛋白；合成；分離；遺傳工程；基因構造；表現。